PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 19 April 15, 1933 Number 4

ON THE DISTRIBUTION OF GALAXIES

By Harlow Shapley
Harvard College Observatory

Communicated March 15, 1933

The Harvard survey of faint external galaxies has now progressed to the point where results on distribution can be presented, based on approximately one hundred thousand systems. The material given in the present summary bears in particular on the problems of the uniformity of distribution of matter in space, the average width of the long-known region of avoidance in low latitude, and the progression with galactic latitude of the number of galaxies per square degree.

The photographic plates used in the survey were all made with the Bruce 24-inch photographic refractor at the Boyden Station of the Harvard Observatory. The earliest plates date from the beginning of the century, but the plates made during the past five years at Bloemfontein by Dr. Paraskevopoulos, Dr. Luyten and Mr. Waterfield have provided the most valuable material. The improvement in the quality and penetration of the plates has been due not only to the skill of the observers, but also to the new mounting constructed for the instrument in 1929 and to the improved photographic emulsions provided by the Cramer Dry Plate Company.

All photographs are on plates fourteen by seventeen inches; the exposures, with few exceptions, are three hours or more in duration. The whole area of each plate was examined for nebulae, but for the statistical investigation the central twenty-five square degrees and the central nine square degrees were separately treated.

For the determination of the magnitude limits of the plates the most satisfactory method is the use of star counts. To establish the plate limits, or the magnitude of any sequence star, the stars in two square degrees are counted. In the present investigation the tables of van Rhijn have been used for regions south of declination $-23°$ to determine the photographic magnitudes from the star counts; the tables by Seares and van Rhijn are used with equal weight for declinations north of $-23°$. For galactic latitudes less than $\pm 20°$ no attempt has been made to de-

termine the limiting magnitudes because of the general inadequacy of the star count method and the absence of satisfactory photographic sequences in the southern sky for stars to the eighteenth magnitude. When the limits of these Bruce photographic plates can be more accurately determined, after further work on magnitude standards and stellar distribution, the data now accumulated should be thoroughly rediscussed. At present we must admit an uncertainty of two- or three-tenths of a magnitude in the plate limits and a corresponding uncertainty in determining the number of galaxies per square degree to a standard magnitude limit.

In the accompanying table the results of the survey are assembled. The first column contains the limits of galactic latitude; the second, the number of fields; the third, the mean values of galactic latitude; and in the following columns are given the mean numbers of galaxies per square degree throughout the corresponding latitude zones.

DISTRIBUTION OF GALAXIES IN GALACTIC LATITUDE

GALACTIC LATITUDE LIMITS	NUMBER OF REGIONS	MEAN GALACTIC LATITUDE	MEAN LOG N	AVERAGE DEVIATION	MEAN LOG N_R	AVERAGE DEVIATION
−90 to −81	6	−83.0	1.26	0.10	0.97	0.16
−80 to −71	10	−76.2	1.37	0.24	1.17	0.37
−70 to −61	12	−64.3	1.29	0.30	1.09	0.26
−60 to −51	26	−54.3	1.09	0.32	1.20	0.41
−50 to −41	41	−45.8	1.05	0.29	1.05	0.36
−40 to −36	15	−37.4	0.78	0.42	0.82	0.42
−35 to −31	21	−32.9	0.87	0.46	1.08	0.57
−30 to −26	20	−28.2	0.62	0.54	0.71	0.52
−25 to −21	28	−23.2	+0.46	0.42	0.56	0.55
−20 to −16	24	−17.7	−0.04	0.55	..	..
−15 to −11	20	−13.1	−0.42	0.49	..	..
−10 to 0	12	− 6.5	−0.28	0.41	..	..
0 to +10	16	+ 6.7	−0.64	0.35	..	..
+11 to +15	11	+13.7	−0.21	0.43	..	..
+16 to +20	10	+18.6	+0.16	0.47	..	..
+21 to +25	6	+22.8	0.81	0.36	1.02	0.18
+26 to +30	6	+27.6	1.23	0.28	1.48	0.33
+31 to +35	13	+33.0	1.02	0.36	1.30	0.37
+36 to +40	11	+37.2	1.19	0.26	1.36	0.23
+41 to +50	18	+45.0	1.28	0.24	1.42	0.25
+51 to +60	10	+54.7	1.20	0.17	1.44	0.24
+61 to +70	11	+65.4	1.21	0.20	1.45	0.14
+71 to +80	8	+74.6	1.27	0.21	1.32	0.23
+81 to +90	3	+85.2	1.30	0.42	1.44	0.73

For this tabulation only the central nine square degrees on each plate have been used. Over such an area there is practically no distance correction for magnitudes of stars or nebulae and on good plates practically no distortion. Plates of low quality have not been used. In only five fields have the clusters of galaxies been conspicuous enough to justify their

omission in the determination of the densities for those regions; and the corrections to the mean values were scarcely appreciable.

In the fourth and fifth columns are given, for each interval of latitude, the mean of the logarithms of the numbers of galaxies per square degree, without correction for plate limits, and the average deviation of a single field from the mean.

The magnitude limits of the plates vary widely because of varying sky conditions and the different speeds of photographic emulsions. For the mean values in columns six and seven we have reduced the number of galaxies in each region to that which would be found to magnitude 18.2,

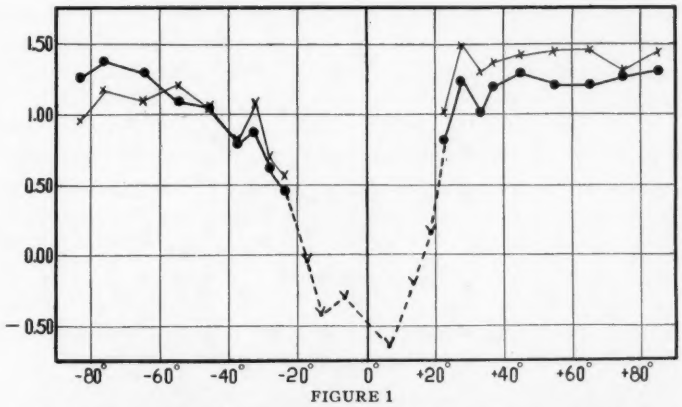


FIGURE 1

Distribution of galaxies in galactic latitude. Coördinates are logarithms of the average number per square degree and galactic latitudes. Values derived directly from counts are indicated by dots; crosses represent values reduced to magnitude limit 18.2. Below $\pm 20°$ the mean points are indefinite because of the total absence of galaxies from some regions.

if we assume that the distribution in space is uniform for each separate region. Thus for plates for which the faintest star magnitude is 17.7 the logarithm of the number of galaxies per square degree has been increased by 0.3, and decreased by the same amount for plates reaching stars of magnitude 18.7. Seventy per cent of all the plates used in the present survey have magnitude limits between these two values. No other corrections to the nebular counts, or adjustments, have been made. The plates made in the rainy season at the Bloemfontein station go less faint, with the result, as shown in figure 1, that in the northern galactic hemisphere the reduction to magnitude 18.2 consistently increases the number of nebulae.

The faintest nebulae shown on any photograph are usually at the plate

limit, but in general the survey on the Bruce plates can be considered as reasonably complete only to a limit about three-tenths of a magnitude brighter than the limit of the plate.

The results tabulated above are shown in figure 1. Some important features are to be noted. The first is the variation of the means with

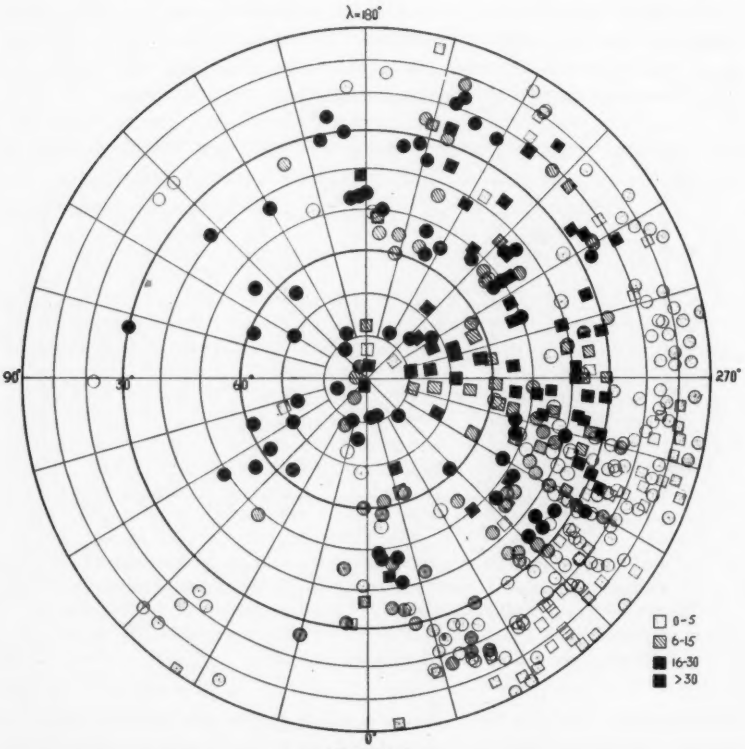


FIGURE 2

Distribution of fields in galactic coördinates. Squares indicate northern latitudes; circles, southern latitudes. Shading represents density per square degree.

galactic latitude, β, a result bearing on the old problem of the concentration of galaxies to the galactic poles. There is no change with latitude north of $+25°$; south of $-25°$ the mean density increases, but the obstructing streamers of dark nebulosity in latitudes $-20°$ to $-40°$ are probably largely responsible for the apparent increase, especially in the reduced values. The avoidance of low galactic latitudes is emphatically indicated.

The second point of interest is the large size of the tabulated average

deviations, in both the original counts and the numbers after reduction to magnitude 18.2. These deviations indicate the conspicuous non-uniformity in the distribution of galaxies throughout the region covered by the survey. The irregularities are so great from field to field that perhaps no conclusion should be drawn concerning the progression of average density with galactic latitude.

The greater richness of the northern galactic hemisphere is as obvious here as in the earlier survey to the thirteenth magnitude. North of latitude $+40°$ the average number per square degree to magnitude 18.2 is 17.8; south of $-40°$ it is 12.6. The difference may be partially the result of our location north of the galactic plane, but the large deviations from region to region show that little trust can be placed in quantitative deductions based on these values.

Figure 2 shows the relation of density to galactic latitude and longitude for the individual regions. An area on the chart is proportional to the corresponding surface area, a result obtained by making the radii of the boundary lines of the various latitude zones proportional to $\sqrt{1 - \sin \beta}$. The densities are the unreduced values, and illustrate again the unevenness in distribution over the sky and the absence of galaxies in regions near the Milky Way.

INFLUENCE OF INTENSIVE DESICCATION ON CERTAIN PHYSICAL PROPERTIES OF BENZENE

BY ALAN W. C. MENZIES AND DONALD A. LACOSS

FRICK CHEMICAL LABORATORY, PRINCETON UNIVERSITY

Communicated March 13, 1933

It has been known for over a hundred years that the presence of traces of water may influence the speed, or the occurrence, of chemical reactions. That such traces of water may likewise affect what are called the physical properties of substances is a possibility that has received attention only in more recent times.

Following a private communication in 1910 of the results, published in the next year, obtained by Smith and Menzies[1] for intensively desiccated calomel, H. B. Baker[2] initiated experiments in which a number of different liquids were intensively desiccated by phosphorus pentoxide. After a period of nine years' drying, Baker reported rises in the boiling points of these liquids ranging from 14 to 60 degrees; and rises also in surface tension and in freezing point. These reports were so extraordinary that many critical investigators were attracted into this field of work. Without attempting to enter into details, it may be stated broadly that the

rises in boiling point reported by Baker have been explained away, in their four several laboratories, by Smits,[3] Smith,[4] Lenher[5] and Cohen with Cohen de Meester[6] in terms of accidental superheating, which occurs also with undried liquids. Failure to observe changes in physical properties as a result of intensive desiccation has been reported by Lenher and Daniels;[7] Briscoe, Peel and Robinson;[8] Rodebush and Michalek;[9] Greer;[10] J. W. Smith;[4] and (although not in the earlier) in the more recent investigations of Smits and his collaborators.[3] The results of Mali[11] and of Manley,[12] on the other hand, are not inconsistent with those of Baker. Upon one matter there appears to be no stated disagreement, namely, that intensive desiccation retards speed of distillation.

In returning to this confused subject after a lapse of twenty years, we thought it best to limit ourselves in the first place to the study of a reputedly normal type of liquid which could readily be purified, and thus we selected benzene. It was only later that it was observed by ourselves[13] that liquid benzene exhibits evidence of allotropic change in the neighborhood of 40°. Because of its sensitiveness, we selected vapor pressure as a chief property to study.

Since we felt that the conditions of experiment were all-important, we were at pains to devise for the work a glass apparatus which, while very simple, enabled us to meet the following specifications: (*a*) superheating cannot occur; (*b*) time at equilibrium can be protracted at will; (*c*) every portion of the glass used can be freed from capillaries; (*d*) all of the completed apparatus can be strongly heated; (*e*) no tubes of capillary bore form part of the apparatus; (*f*) no glass septum need be broken within the dry apparatus; (*g*) no mercury or stopcocks are used; (*h*) the purified phosphorus pentoxide and purified benzene can be redistilled direct into the apparatus; (*i*) the mass of desiccant used can be many times that of the substance to be dried; (*j*) the apparatus can be charged under dust-free conditions; (*k*) desiccation at elevated temperatures is possible.

The importance of conditions (*d*) and (*j*) has been pointed out elsewhere.[14]

The apparatus consisted of a simplified isoteniscope,[15] with a 2-cc. bulb *A*, connected as shown in the diagram to a cylindrical bulb *B* of about 70-cc. volume, which contained the desiccant. During vapor pressure measurements, the latter bulb could be jacketed by water of

known temperature, while the small bulb, containing benzene, was immersed in a stirred oil bath, according to standard isoteniscope technique. The small tubes, shown sealed off, at *C*, *D* and *E* were used for introducing the materials and for currents of dust-free oxygen or nitrogen. The apparatus could finally be sealed off with a content of many grams of phosphorus pentoxide, about one gram of solidified benzene at $-75°$, and nitrogen gas at known temperature and pressure, near that of the room.

The temperature at which the vapor pressure of the liquid benzene was equal to that of the nitrogen, with its partial pressure of benzene vapor, at 18° may be called the boiling point of the sample. After expulsion of the nitrogen from bulb *A* by boiling out, the identity of these two pressures is recognized by the equality of level of the liquid benzene surfaces in the two limbs of the trap *F*. The boiling point determined in this way immediately after sealing coincided with the known boiling point of the sample as previously ascertained by the customary methods. It has been shown elsewhere[16] that the water content of such benzene is insufficient to cause by its removal a boiling point change near 80° of over 0.1°.

For purposes of desiccation, the liquid benzene was transferred to bulb *B* in contact with the desiccant. When, after a lapse of weeks or months, it was desired to redetermine the boiling point, the benzene was distilled at room temperature into bulb *A*, maintained at a lower temperature. In most of our experiments we permitted the drying process to take place at room temperature. In other experiments, we used temperatures as high as 150°; in which cases, to minimize danger of bursting the apparatus, we enclosed it completely in an iron tube containing benzene whose vapor pressure at the elevated temperature offset that of the benzene within the glass apparatus.

Some of the experimental results which we obtained by these methods may be summarized as follows:

1. When benzene was dried with ordinary phosphorus pentoxide no change in its vapor pressure was observed.

2. When benzene was dried with purified phosphorus pentoxide but without precautions to exclude dust, no change in its vapor pressure was observed.

3. When benzene was dried with purified phosphorus pentoxide with precautions to exclude or destroy dust, its vapor pressure near 80° was (*a*) markedly lowered if the drying process has been carried on at room temperatures, and (*b*) slightly raised if the drying process had been carried on above 80°.

4. In certain cases, under the conditions of (3), no change was observed, a result attributable, for example, to accidental dust, which likewise is sufficient to explain the negative results obtained by others in desiccation experiments.

5. Interpreted in terms of change of boiling point near atmospheric pressure, desiccation under our conditions at room temperature for two months caused a rise of about 0.4°, for four months about 0.7°, while the maximum rise observed, practically stationary after twenty months, was 2.2°.

6. No evidence was found of "fractionation," even in a sample of benzene whose boiling point had been raised 2.2° by desiccation.

7. When ordinary undried air was slowly admitted by diffusion to benzene whose vapor pressure had been reduced by desiccation to an abnormally low value, the vapor pressure reverted again to the normal value in a few hours.

8. In regard to reversion to normal boiling point while within the sealed apparatus, no sample of benzene showed such reversion after standing at room temperature. Two samples whose boiling points had been raised 0.4° by two months' drying at room temperature showed complete reversion upon heating near 78° out of direct contact with phosphorus pentoxide for at most three days; whereas a sample whose boiling point had been raised 2.2° by eighteen months' drying at room temperature showed no change of boiling point in either direction upon heating for periods of 18, 4 and 9 days at 120°, 135° and 150°, respectively.

9. Samples of desiccated benzene of abnormally high boiling point showed no change from the normal freezing point greater than 0.1°.

[1] Smith, A., and Menzies, A. W. C., *Z. physik. Chem.*, **76,** 713 (1911).

[2] Baker, H. B., and Baker, M., *J. Chem. Soc.*, **101,** 2339 (1912); Baker, H. B., *Ibid.*, **121,** 568 (1922); **123,** 1223 (1923); 2902 (1927); 1051 (1928); 1661 (1929).

[3] Smits, A. (and collaborators), *Z. physik. Chem.*, **76,** 421 (1911); **100,** 477 (1922); **129,** 33 (1927); *J. Chem. Soc.*, **125,** 1069 (1924); **125,** 2554, 2573 (1924); 2657 (1926); 2399 (1928); 2712, 2724 (1929); *Verslag. Akad. Wetenschaffen Amsterdam,* **34,** 468 (1931).

[4] Smith, J. W., *J. Chem. Soc.*, 867 (1927); 788 (1929); 528, 2573 (1931).

[5] Lenher, S., *J. Phys. Chem.*, **33,** 1579 (1929).

[6] Cohen, E., and Cohen de Meester, *Verslag. Akad. Wetenschaffen Amsterdam,* **33,** 1003 (1930).

[7] Lenher, S., and Daniels, F., *Proc. Nat. Acad. Sci.*, **14,** 607 (1929).

[8] Briscoe, H. V. A., Peel, J. B., and Robinson, P. L., *J. Chem. Soc.*, 368 (1929).

[9] Rodebush, W. H., and Michalek, J. C., *J. Am. Chem. Soc.*, **51,** 748 (1929).

[10] Greer, E. J., *Ibid.*, **52,** 4191 (1930).

[11] Mali, S. B., *Z. anorg. Allgem. Chem.*, **149,** 150 (1925).

[12] Manley, J. J., *Phil. Mag.*, **83,** 249 (1932).

[13] Menzies, A. W. C., and Lacoss, D. A., *Proc. Nat. Acad. Sci.*, **18,** 144 (1932).

[14] Menzies, A. W. C., *Nature,* **125,** 446 (1930); Baker, H. B., *Ibid.*, **125,** 229 (1930).

[15] Smith, A., and Menzies, A. W. C., *J. Am. Chem. Soc.*, **32,** 1412 (1910).

[16] Menzies, A. W. C., *J. Phys. Chem.*, **35,** 1655 (1931).

ON THE STERILITY OF THE INTERRACIAL HYBRIDS IN DROSOPHILA PSEUDOOBSCURA

BY TH. DOBZHANSKY

WILLIAM G. KERCKHOFF LABORATORIES OF THE BIOLOGICAL SCIENCES, CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA

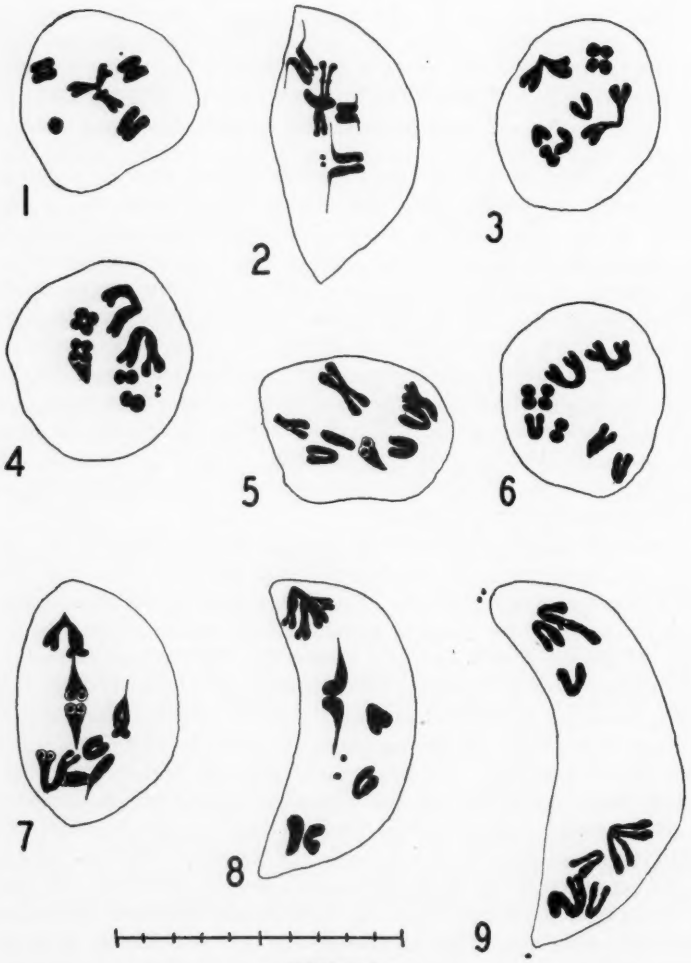
Communicated March 13, 1933

Offspring from interspecific crosses are frequently equal or superior to their parents in somatic vigor, and are, nevertheless, partially or completely sterile. The sterility is due to a disturbance in the process of gametogenesis usually involving a more or less complete lack of chromosome pairing at meiosis. The known instances of interspecific sterility fall into two groups. To the first group belong cases described by Karpechenko,[1] Clausen,[2] Tschermak and Bleier[3] and many others, who have found that the doubling of the chromosome complement of sterile hybrids results in the appearance of fertile allotetraploids showing a more or less complete chromosome pairing. To the second group belongs the case of the interracial hybrids in Drosophila pseudoobscura. As shown below, the doubling of the chromosomes does not influence here either the lack of chromosome pairing or sterility.

Lancefield[4] discovered that the species Drosophila psuedoobscura Frol.[5] is divided into two races or "physiological species." The two races, called race *A* and race *B* respectively, are indistinguishable in appearance. Both races possess five pairs of chromosomes, but the *Y*-chromosome is *J*-shaped in race *A* and *V*-shaped in race *B*. The progeny of the interracial crosses consists of partially fertile females and completely sterile males.

The writer has found that the sterility of the hybrid males is due to a profound disturbance of spermatogenesis, including failure of chromosome pairing at meiosis. The upper part of the testis in the $(A\,♀ \times B\,♂)F_1$ hybrid males[6] consists of spermatogonia and young first spermatocytes which are rather similar to those found in the corresponding part of the testis in normal males. At diakinesis the differences become apparent. In normal males all the chromosomes are paired at diakinesis, and appear as bivalents of the tetrad type (Fig. 1). Usually only four bivalents are observable since one of the autosomes is invisible at meiosis on account of its small size. The other autosomes pair with their homologues along their entire lengths, while the *X*- and *Y*-chromosomes are, as first shown by Metz,[7] paired only for a rather short distance in the region of the spindle fibre attachments.

In hybrid males chromosomes are mostly unpaired at diakinesis, although a variable number of bivalents is present in some cells (Figs. 3–6). Whether a given chromosome visible at diakinesis represents a univalent -



FIGURES 1–9

Figure 1, diakinesis in a normal male (race *A*, four bivalents and the nucleolus); Figure 2, first division in a normal male (race *A*); Figures 3–6, diakinesis in hybrid males; Figures 7–9, first division in hybrid males. The scale for figures 1–17 represents 10 micra.

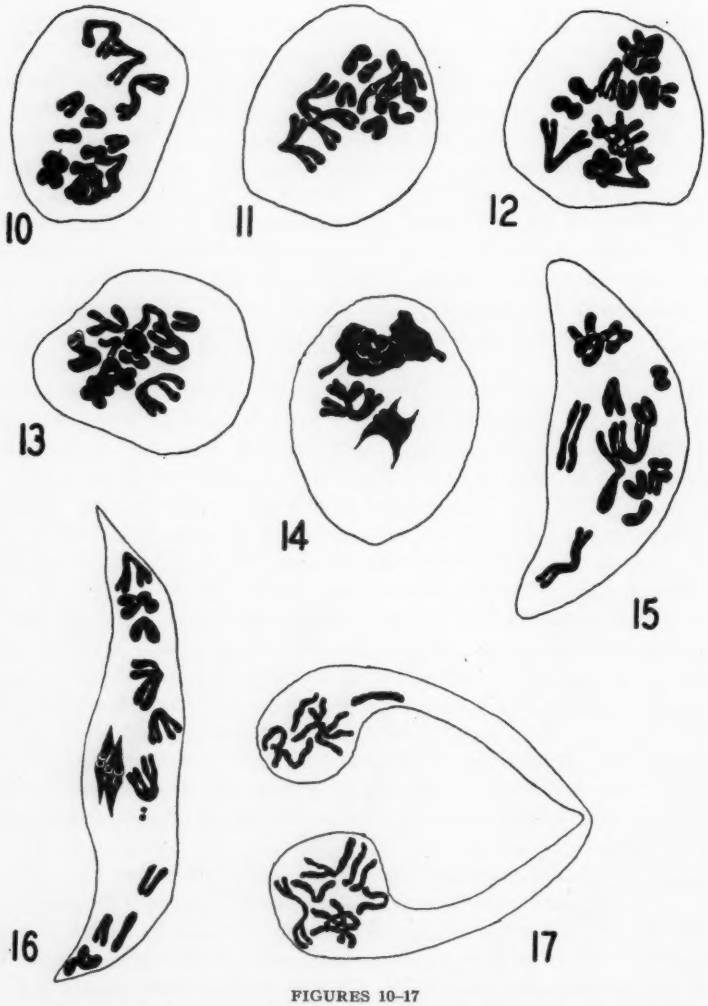
or a bivalent usually can be determined with accuracy. An autosomal bivalent seen in a side view appears as two closely paired rods, the space between which is, however, clearly visible (Fig. 1). In top view autosomal bivalents reveal their tetrad structure (one bivalent in the upper part of Fig. 3, two bivalents in the left part of Fig. 4). The autosomal univalents are split equationally, and appear as dyads. The sister chromatids diverge at their free ends, but are held together at the spindle fibre, which in these autosomes is terminal. Consequently, autosomal univalents appear in side view as more or less short *V*'s (Figs. 3, 5, 6), and in top view as dumb-bell-shaped bodies (lower part of Figs. 3 and 4, left part of Fig. 6). In the *X*- and the *Y*-chromosomes the spindle fibre attachments are median or submedian. Consequently, the *X*-*Y* bivalents appear as cross-shaped or branched bodies (Figs. 1 and 2). The unpaired *X*'s and *Y*'s are *V*-shaped chromosomes both limbs of which show the equational split (Figs. 3–6).

The first meiotic division in normal males has been described by Metz.[7] At anaphase the autosomal bivalents separate first at their spindle fibre ends, the opposite ends remaining associated until late anaphase (Fig. 2). Mitochondria assemble on the surface of the spindle making its outline very clearly visible. The telophase proceeds normally, two second spermatocytes are formed, whose nuclei, after a rather brief interphase during which the chromosomes do not pass into a resting stage, undergo the second division, giving rise to young spermatids.

In hybrid males the first division is highly abnormal. The univalents form no equatorial plate but are scattered on the spindle (Fig. 7). At anaphase the univalents move toward the poles, but their distribution is frequently unequal, so that more chromosomes go to one pole than to the other (Fig. 7). The homologous univalents frequently go to the same pole (see the *X*- and the *Y*-chromosomes at the same pole in Figs. 8 and 9). The bivalents, if any are present in the nucleus, stay on the equator of the spindle longer than do the univalents (Figs. 7 and 8), and disjoin normally.

Already at early anaphase the spindle in hybrids shows signs of an abnormal elongation. Soon the poles of the spindle approach the periphery of the cell, and, since the elongation continues, the spindle begins to bend (Fig. 9). Further enormous elongation of the spindle causes the poles to slide toward each other, the spindle becoming a nearly closed ring (Fig. 17). The daughter nuclei are then formed, but the cytoplasm fails to divide. The second division is absent. The binucleate spermatids transform into giant worm-like bodies somewhat resembling the atypical spermatozoa of certain Prosobranchia. This behavior of the spindle is interesting in connection with Belar's[8] views on the mechanism of mitosis.

Islands of tetraploid tissue are not infrequently met with in the testes



FIGURES 10–17

Figures 10–13, diakinesis in tetraploid spermatocytes in hybrid males; Figure 14, same, early anaphase; Figures 15–16, same, late anaphase; Figure 17, late telophase.

of hybrid males (in about one out of every four testes examined), but so far have not been seen in the testes of normal males. The origin of the tetraploid tissue is due to spermatogonial mitoses not followed by fission of the cytoplasm. Binucleate spermatogonia are thus formed. The two nuclei in such spermatogonia divide again, the two mitotic spindles lying side by side in the same cell. It is, presumably, at the telophase of such divisions that the pairs of daughter nuclei fuse with each other with the consequent formation of tetraploid spermatogonia. The tetraploid spermatogonia, recognizable by their large size, divide further and transform into tetraploid spermatocytes. Meiosis can be observed in the tetraploid spermatocytes.

Figures 10–13 show diakinesis in tetraploid cells. On account of the large number of chromosomes the figures are not as clear as the corresponding figures in the diploid cells. It is, however, easy to see that most chromosomes are unpaired and appear as univalents (dyads). In each of the figures three or four *V*-shaped chromosomes showing the equational split in one or in both arms are present. These chromosomes are unquestionably the unpaired *X*'s and *Y*-chromosomes. In each of the figures one can see also a large number of dyads which are the autosomal univalents. Like the diploid cells, the tetraploid spermatocytes show a varying number of bivalents. In figure 10 at least one bivalent is visible (in the lower left); no bivalents are evident in figure 11; in figure 12 there is at least one bivalent (see the lower part of the figure, under the left arm of the *V*-shaped chromosome); in figure 13 at least two bivalents are present. Some of the figures observed even suggest the possibility of the formation of multivalents. Figure 14 represents an early anaphase. This figure is, on the whole, unclear, but a distinct bivalent may be seen in its upper part, and in its lower part one sees a body with four spindle fibre attachments which are directed two by two in the opposite directions. This body is best interpreted as a quadrivalent (octad).

The course of the first meiotic division in the tetraploid spermatocytes is similar to that observed in the diploid ones. Figure 15 represents an early anaphase. The univalent chromosomes are scattered on the spindle; no bivalents seem to be present. Figure 16 shows a late anaphase. The univalents are going toward the poles, the spindle begins to bend. In the region of the equator of the spindle two clear autosomal bivalents are visible; their spindle fibre ends have already disjoined, the other ends are still associated.

The telophase of the first division in the tetraploid cells shows the same enormous elongation of the spindle which has been observed in the diploid cells. The poles of the spindle converge, and giant binucleate cells are formed. The second division being absent, the binucleate cells transform directly into abnormal octoploid worm-like spermatids.

The doubling of the chromosome complement in the interracial hybrids in Drosophila pseudoobscura does not lead to an increase in the frequency of chromosome pairing at meiosis, nor to restoration of fertility. These hybrids are, therefore, fundamentally different from the allotetraploids described in many species of plants.

One may surmise that the failure of chromosome pairing and the consequent sterility observed in interspecific hybrids may be due to either of two different causes. First, rearrangements of the chromosomal material, such as translocations and inversions, may lead to the chromosomes of one species or race finding no complete homologues among the chromosomes of another species or race. The consequent conflict of the attraction forces (Dobzhansky,[9] Darlington[10]) results in a failure of chromosome pairing in the hybrids. The doubling of the chromosome complement furnishes an exact homologue to every chromosome. Chromosomes pair normally at meiosis, and the sterility is eliminated. Second, the failure of chromosome pairing in the hybrids may be due to the action of complementary genetic factors contributed by the parent species or races involved in the cross. In this case the doubling of the chromosome number should produce no change in the course of meiosis, and the hybrid sterility should be preserved.

The sterility of the hybrids giving fertile allotetraploids is probably due to the factors of the first type. The factors of the second type seem to be responsible for the sterility of the interracial hybrids in Drosophila pseudoobscura (cf. Dobzhansky and Boche[11]), and probably also for the sterility of the hybrids between Drosophila melanogaster and Drosophila simulans (Schultz and Dobzhansky[12]).

[1] Karpechenko, G. D., "The Production of Polyploid Gametes in Hybrids," *Hereditas*, **9**, 349–368 (1927*a*); Karpechenko, G. D., "Polyploid Hybrids of Raphanus sativus L. Brassica oleracea L.," *Bull. Appl. Botany*, **17**, 305–410 (1927*b*).

[2] Clausen, R. E., "Interspecific Hybridization in Nizotiana. 7. The Cytology of Hybrids of the Synthetic Species, Digluta, with Its Parents, Glutinosa and Tabacum," *Univ. Calif. Publ. Botany*, **11**, 177–211 (1928).

[3] Tschermak, E., and Bleier, H., "Über fruchtbare Aegilops-Weizenbastarde," *Ber. Deut. Botan. Ges.*, **48**, 110–132 (1926).

[4] Lancefield, D. E., "A Genetic Study of Two Races or Physiological Species in Drosophila Obscura," *Zeit. ind. Abst. Vererbgsl.*, **52**, 287–317 (1929).

[5] The American species Drosophila pseudoobscura Frol. is frequently confused with the European species Drosophila obscura Fall. Frolova and Astaurow (*Zeit. Zellf. mikr. Anat.*, **10** (1929)) have shown that these two forms are unquestionably specifically different.

[6] The present paper is devoted to the description of the spermatogenesis in the $(A\,♀ \times B\,♂)F_1$ hybrid males only. Spermatogenesis in the $(B\,♀ \times A\,♂)F_1$ hybrid males is different in many important details, and will be described elsewhere. In both cases the failure of chromosome pairing at meiosis is, however, observed.

[7] Metz, C. W., "Observations on Spermatogenesis in Drosophila," *Zeit. Zellf. mikr. Anat.*, **4**, 1–28 (1926).

[8] Belar, K., "Beiträge zur Kausalanalyse der Mitose. 2. Untersuchungen an den Spermatocyten von Chorthippus (Stenobotrus) lineatus Panz," *Roux Arch. Entwmech.*, **118**, 359–484 (1929).

[9] Dobzhansky, Th., "The Decrease of Crossing-Over Observed in Translocations, and Its Probable Explanation," *Amer. Natur.*, **65**, 214–232 (1931); Dobzhansky, Th., "Studies on Chromosome Conjugation. I. Translocations Involving the Second and the *Y*-Chromosomes of Drosophila Melanogaster," *Zeit. ind. Abst. Vererb.*, **60**, 235–286 (1932).

[10] Darlington, C. D., *Recent Advances in Cytology*, Philadelphia, Blakiston's, 559 pp. (1932).

[11] Dobzhansky, Th., and Boche, R. D., "Intersterile Races of Drosophila pseudoobscura," *Biol. Zbl.*, in press.

[12] Schultz, J., and Dobzhansky, Th., "Triploid Hybrids between Drosophila melanogaster and Drosophila simulans," *J. Exptl. Zoöl.*, in press.

A FURTHER STUDY OF BLOOD GROUPS OF THE RABBIT

BY CLYDE E. KEELER AND W. E. CASTLE

HOWE LABORATORY AND BUSSEY INSTITUTION, HARVARD UNIVERSITY

Communicated March 6, 1933

It has been shown in previous publications that, by means of immune sera, domestic rabbits may be separated into four groups characterized by the presence or absence from the blood of two different hæmagglutinogens, which we have called H_1 and H_2.[1] These agglutinogens are inherited as Mendelian dominant characters. The constitution of the four groups as regards the agglutinogens is as follows. Rabbits of one group (the double dominant) possess both agglutinogens. We may refer to them as H_1H_2 animals. Rabbits of a second group (the double recessive) possess neither agglutinogen. They may be called the *O* group. Rabbits of the other two groups are single dominant animals, possessing, respectively, either H_1 or H_2 alone. The genes for H_1 and H_2 are borne on the same chromosome, an unidentified autosome, since the inheritance is not sex-linked. These two genes are either allelomorphs or closely linked, more probably allelomorphs.

We are able to confirm the generally accepted view that such hæmagglutinogens as these are borne in or on the red blood-corpuscles. For upon injection of *washed* blood-corpuscles from a rabbit possessing an agglutinogen into the body of a rabbit *not* possessing it, a specific antibody (agglutinin) for that same agglutinogen makes its appearance in the serum of the recipient. Thus (*a*) if blood from an H_1 rabbit is injected repeatedly into an H_2 rabbit or a *O* rabbit, the serum of the recipient will later be found to contain an antibody (agglutinin) which will

cause agglutination of H_1 blood-corpuscles, but will not cause agglutination in the blood of an H_2 or O individual. Also (*b*) if blood from an H_2 rabbit is injected into either an H_1 or a O rabbit, the recipient will develop a specific H_2 antibody (agglutinin) which will clump H_2 blood-corpuscles but will not affect the corpuscles of H_1 or O animals. And (*c*) if blood from an H_1H_2 rabbit is injected into a O rabbit, two antibodies are simultaneously produced, which are the specific antibodies for H_1 and H_2, respectively. For serum thus produced will agglutinate blood-corpuscles from H_1 animals, H_2 animals and H_1H_2 animals. It fails to cause agglutination only in the case of the blood from O animals.

A question of practical importance arose as to how early in the life of the individual rabbit the agglutinogens are present in effective form. Could we, for example, ascertain with certainty in new-born rabbits the presence or absence from the blood of a particular agglutinogen? If so, our inheritance studies would be greatly expedited. Serological tests made upon the blood of new-born rabbits gave results as emphatically plus or minus as in the case of older individuals. Also the qualitative results were in every case in harmony with expectation based on tests made on older individuals. Accordingly, we concluded that the agglutinogens are present in full strength in new-born individuals. Our experiments indicate clearly that they do not subsequently change.

We next extended the inquiry to prenatal stages of development. For this purpose, an H_1 female was killed fifteen days after she had been mated to a double dominant (H_1H_2) male. She was found to contain eight embryos averaging about 14 mm. in length. It was impossible by ordinary procedures to obtain from each embryo separately enough blood for a serological test. So the entire embryo was pulped in citrate solution and centrifuged, after which as much tissue as possible was removed mechanically. The fluid thus obtained, which was of a pinkish color from the contained red blood-cells, was subjected to the usual tests for presence of the agglutinogens. It should be stated that the red blood-cells in these embryos were large, oval and nucleated. All the embryos gave positive reactions for presence of one or both agglutinogens, as was to be expected from the genetic constitution of the parents.

The mother, an H_1 individual, was heterozygous for the H_1 gene and so would transmit in her gametes either H_1 or O. The father, an H_1H_2 individual, would be heterozygous for both genes and so would transmit either H_1 or H_2 in every gamete. Consequently the genetic set-up would be

Mother's gametes	H_1 or O
Father's gametes	H_1 or H_2
Expected combinations	H_1H_1, H_1, H_1H_2 and H_2

Of the four expected combinations, all different, one would be homo-

zygous H_1, another would be heterozygous H_1 (both giving the same serological reaction), a third would be H_1H_2 (heterozygous for both) and the fourth a heterozygous H_2. Qualitatively, the tests would be expected to show three kinds of individuals in the ratio $2H_1:1H_1H_2:1H_2$. Actually the numbers found were $2H_1:5H_1H_2:1H_2$. Theoretically no O individuals could result from this mating, and in fact none were found. The agglutination tests were in the case of every embryo clear cut showing unmistakable clumping of the red blood-cells as stated.

We conclude that the agglutinogens are probably developed as early as the red blood-cells themselves, certainly as early as the fifteenth day of embryonic development.

A question arose as to whether the agglutinogens might not be present in tissues other than the red blood-cells as, for example, in lymph cells. But no evidence could be obtained that such is the case. We conclude that the current view is probably correct that the haemagglutinogens are developed only in or on the red blood-cells.

The situation is very different as regards the agglutinins. These are antibodies developed in the blood-serum in reaction to foreign agglutinogens artificially introduced into the body. A rabbit does not develop an antibody against an agglutinogen which it possesses. If it did, its own blood-stream would become blocked.

An H_1 rabbit is incapable of developing an H_1 antibody (agglutinin) but will develop H_2 agglutinin if H_2 blood is introduced into its body. Also an H_2 rabbit is incapable of developing H_2 agglutinin, but will develop H_1 agglutinin if H_1 blood is introduced into its body. An H_1H_2 rabbit can develop neither agglutinin, since it already possesses both agglutinogens; but a O rabbit can develop both H_1 and H_2 agglutinins, if blood containing both agglutinogens is received into its body.

A mixture of sera containing, respectively, H_1 and H_2 agglutinins has the same serological properties as the serum of a O rabbit which has received injections of H_1H_2 blood. This shows that the formation of H_1 agglutinin is not interfered with by the formation of H_2 agglutinin simultaneously in the same serum.

Agglutinogens do not pass through the placenta, since embryos may, as in the case already described, possess an agglutinogen not possessed by the mother, yet without causing an agglutinin to form in her blood. Conversely, in the same experiment it is shown that embryos may lack an agglutinogen possessed by the mother. Hence agglutinogens do not pass freely through the placenta. It is known also that blood-corpuscles do not pass through the placenta, and this strengthens the conclusion that the agglutinogens are borne by the red blood-corpuscles.

Agglutinins (antibodies formed in the blood-serum and so not attached to blood-corpuscles) *do* pass through the placenta, as is shown by the

following experiment. A zero female rabbit (possessing neither agglutinogen) was mated to a male of the same constitution, a mating which could only result in the production of young also *O* (lacking both agglutinogens). After allowing sufficient time for the foetuses to implant (about 7 or 8 days) the doe was given at 2-day intervals 6 intraperitoneal injections of blood from an H_1H_2 rabbit (possessed of both agglutinogens). On the last day of pregnancy the blood-serum of the mother was tested and found to have developed strong agglutinins which clumped promptly blood-cells of the donor.

Seven young were born from this mating, and their blood, when tested against both H_1 serum and H_2 serum, was found, as expected, to be devoid of agglutinogens H_1 and H_2, placing them in the *O* group. Their serum, however, was found in every case to cause agglutination of both H_1 and H_2 blood-cells. Hence, both agglutinins, developed in the blood-serum of the mother in reaction to introduced foreign H_1H_2 blood, had been able to pass through the placenta into the blood-serum of each of the seven young. The only possible alternative explanation would be to suppose that the agglutinogens themselves, introduced in foreign blood into the body of the mother, had passed through the placenta into the embryos, which had then developed agglutinins as the mother had already done. This seems improbable for reasons already stated in support of the idea that agglutinogens are borne in or on the red blood-cells, which are unable to pass through the placenta. Also the possibility of agglutinin development within embryos is rendered improbable by the following experiment. An H_1 rabbit, on the day of its birth and daily thereafter for four more days, was given an injection of H_1H_2 blood. Subsequently for several weeks bi-weekly injections were given without the production of H_2 agglutinin in its serum, a result which would have been expected promptly in an adult rabbit of the same blood group. We conclude that agglutinins are not developed readily in very young rabbits, whereas agglutinogens are present even in embryonic stages.

Agglutination of the red blood-cells occurs in the rabbit, so far as known observations go, only when an agglutinogen, borne by the blood-cells, meets with an agglutinin developed in the serum of a rabbit which *lacks* that agglutinogen in its blood-cells, but has had the agglutinogen introduced into its body in foreign blood.

Why an animal accumulates no antibody to its own agglutinogen or agglutinogens remains obscure. Certainly such an occurrence would be disastrous, but that affords no explanation of its failure to occur. Possibly agglutinin *is* formed continuously in the presence of an agglutinogen continuously present, and is as continuously satisfied or neutralized so that it has no chance to accumulate in harmful quantities. Whereas when a foreign agglutinogen is introduced, an antibody is produced, to

neutralize or absorb which there is present no continuing supply of agglutinogen.

Be this as it may, the rule is obvious enough, that an agglutinin forms in the serum only when a foreign agglutinogen (one not already present in the blood of the individual) is introduced repeatedly into its body. This rule can be tested in sixteen different kinds of blood transfers involving individuals of the four blood groups, as shown in table 1.

TABLE 1

POSSIBLE BLOOD TRANSFERS WITHIN THE FOUR BLOOD GROUPS OF THE RABBIT, WITH THE EXPECTED RESULTS AS REGARDS PRODUCTION OF AGGLUTININS

CASE NO.	GROUP OF DONOR	GROUP OF RECIPIENT	AGGLUTININ FORMATION EXPECTED	TIMES TESTED AND VERIFIED
1	O	O	None	2
2	O	H_1	None	0
3	O	H_2	None	0
4	O	H_1H_2	None	0
5	H_1	O	H_1	0
6	H_1	H_1	None	0
7	H_1	H_2	H_1	1
8	H_1	H_1H_2	None	0
9	H_2	O	H_2	1
10	H_2	H_1	H_2	1
11	H_2	H_2	None	0
12	H_2	H_1H_2	None	0
13	H_1H_2	O	H_1 and H_2	2
14	H_1H_2	H_1	H_2	0
15	H_1H_2	H_2	H_1	1
16	H_1H_2	H_1H_2	None	4

We have actually made seven of the sixteen possible transfers indicated in table 1, and in every instance have secured serum of the expected character as regards presence or absence of agglutinins. The experiment has been repeated twice in Case No. 1 and also in Case No. 13, and four times in Case No. 16.

We have attempted to determine more precisely the portion of the red blood-cells which carries the agglutinogens but thus far with negative results.

First a mixture was made of H_1 blood-cells with H_2 blood and this was then centrifuged at a high rate of speed to force the cells into intimate contact. The cells were then shaken up and treated with H_2 serum, which caused part of the cells to become agglutinated. The remaining free cells were then pipetted off and treated with H_1 serum, which promptly caused their agglutination. Hence it seems probable that no change had occurred in the agglutinogen load of the cells composing the mixture. Part of them, as originally, bore the H_1 agglutinogen, and the remainder

bore the H_2 agglutinogen. Attempts to transfer agglutinogen to O cells by centrifuging were equally futile.

Next, blood-cells of an H_2 rabbit were burst by treatment with distilled water, and the cell contents were centrifuged. But from the separated cell contents, in a ring-test with H_2 serum, no reaction was obtained. Hence it seems probable that intact erythrocytes are essential to the reactions.

Relative strengths of the H_1 and H_2 reactions have been tested in the following way. A test serum was made by injecting blood from an H_1H_2 animal into a O animal. This bivalent serum was, as expected, able to agglutinate blood-cells of all groups of rabbits except zero animals. A litter of twelve new-born young produced by a zero mother mated with an H_1H_2 male was now subjected to comparative tests with H_1 serum, H_2 serum and the before-mentioned H_1H_2 serum. Each individual of the litter was, as expected from the nature of the mating, positive to *only one* of the univalent sera and negative to the other. Seven individuals were positive to the H_2 serum, and five were positive to the H_1 serum. These two groups of individuals differed consistently in the strength of their reaction to the bivalent serum. The seven H_2 individuals gave a stronger reaction (indicated as ++++ in our records) whereas the five H_1 individuals gave a reaction indicated as + merely in our records. We conclude that the H_2 reaction, other things being equal, is stronger than the H_1 reaction.

Comparative tests between homozygous and heterozygous H_1 animals (as well as between homozygous and heterozygous H_2 individuals) revealed no difference in the strength of the reaction, indicating that so far as this test is concerned, dominance is complete, one genetic dose of an agglutinogen being as effective as two in producing agglutination.

Considerable attention has been given to the question whether a mixture of the univalent H_1 and H_2 sera has the same physiological properties as the bivalent H_1H_2 serum obtained when blood of a double dominant (H_1H_2) animal is injected into a O animal. So far as we can discover, their properties are identical. This is indicated by the following observations.

(1) A mixture of H_1 and H_2 sera completely agglutinates H_1 blood, H_2 blood and H_1H_2 blood. The bivalent H_1H_2 serum does the same. Neither the mixed univalent sera nor the bivalent serum agglutinates O blood. (2) When a mixture is made of H_1 blood and H_2 blood, and this is treated with H_1 serum, part only of the blood-cells are agglutinated. But addition of H_2 serum will cause the remaining blood-cells to become agglutinated. If a similar mixture of H_1 and H_2 blood is treated with the bivalent H_1H_2 serum, all blood-cells are simultaneously agglutinated.

These facts indicate that the two agglutination reactions occur inde-

pendently. A specific agglutinogen and a specific agglutinin are necessary to each. When these are present together, both reactions occur simultaneously. But if either agglutinogen or either agglutinin is withheld, the corresponding reaction is lacking.

Absorption tests failed to show any diminution of the potency of either univalent serum by the presence of the other.

The question has sometimes been raised, in discussions of blood group inheritance, whether agglutinins are inherited independently of agglutinogens. In the case of the rabbit, no uncertainty can exist. The only thing which is inherited, as a basis for blood-group formation, is the agglutinogen. An agglutinin arises only as an antibody developed in the blood-serum in response to introduced foreign blood carrying an agglutinogen not present in the blood of the recipient. It is not an inherited but an acquired property of the individual.

In the case of the *A-B* blood groups of man, the agglutinins are not acquired but natural properties of the blood of individuals. Their existence would seem to antedate phylogenetically that of the agglutinogens, and where they fail to be found, it would seem to be due to suppression in consequence of the appearance of a complementary agglutinogen. Thus we find that specific antibodies (agglutinins) for the *A* and *B* agglutinogens are universally present in human blood except where a corresponding agglutinogen has made its appearance. There the agglutinin is wanting.

Thus, *O* individuals have both *a* and *b* agglutinins;
A individuals have only *b* agglutinin;
B individuals have only *a* agglutinin;
and *AB* individuals have neither *a* nor *b* agglutinin.

Perhaps the agglutinins are natural constituents of human blood-serum, having some other physiological function than that known to us, and caused to disappear only upon the appearance of a complementary agglutinogen, the continuous presence of which absorbs the agglutinin as fast as it is formed.

In that case we should be justified in considering the agglutinins not as special, gene-determined structures, any more than cytoplasm is so considered in relation to the cell. Agglutinogens, on the other hand, are clearly gene-determined structures, subject to mutation, having come into existence doubtless by that method, and now clearly inherited as other mutational differences are, in accordance with Mendel's law.

It seems, accordingly, futile to speak of agglutinins as inherited structures either in the rabbit or in man. What is inherited is clearly the agglutinogen in both cases. The agglutinin is merely a correlated structure dependent for its *presence* upon the *absence* of a related agglutinogen to absorb it.

Summary.—1. The four blood groups of the rabbit, demonstrable by means of immune sera, depend upon the presence or absence of two different agglutinogens, which are borne by the red blood-cells.

2. These agglutinogens (designated H_1 and H_2, respectively) are inherited as Mendelian dominant characters, and behave as allelomorphs (or closely linked genes) in transmission.

3. The agglutinogens are present in effective form in new-born rabbits. They are present even earlier in embryos 15 days old having nucleolated erythrocytes.

4. Agglutinogens can be detected in no cells other than erythrocytes.

5. Agglutination occurs only when an agglutinogen meets its specific antibody known as an agglutinin.

6. An agglutinogen and its antibody are not normally present in the same individual.

7. The antibody (agglutinin) is produced in the blood-serum of a rabbit which has received injections of blood containing an agglutinogen foreign to the recipient.

8. Agglutinogens do not pass through the placenta, neither do erythrocytes, the supposed bearers of the agglutinogens.

9. Agglutinins, which are not attached to the blood cells but are borne in the blood-serum, *do* pass through the placenta.

10. An explanation of the failure of a rabbit to develop an agglutinin against its own agglutinogen is perhaps found in the capacity of the ever present agglutinogen to absorb or neutralize the agglutinin as fast as it is produced.

11. Attempts to discover the localization of agglutinogens on or in the erythrocytes have been fruitless.

12. An experiment with bivalent H_1H_2 serum in comparison with univalent H_1 and H_2 sera indicates that the H_2 reaction is stronger than the H_1 reaction.

13. No difference is discoverable between the strength of the reaction in an individual heterozygous for an agglutinogen as compared with that in a homozygous individual.

14. A mixture of univalent H_1 and H_2 sera has the same physiological properties as bivalent H_1H_2 serum. Neither univalent serum interferes with the action of the other.

15. Agglutinogens are clearly gene-determined mutational characters inherited in accordance with Mendel's law; agglutinins are not to be regarded as independently inherited characters, but as conditions correlated with presence or absence of certain agglutinogens. The question of inheritance concerns only the agglutinogens.

[1] For purposes of comparison it may be noted that our H_1 corresponds with B of Fischer and Klinkhardt, and our H_2 with the A of those authors.

Castle, W. E., and Keeler, C. E., *Proc. Nat. Acad. Sci.*, **19**, 92–100 (1933).
Fischer, W., and Klinkhardt, G., *Arb. Staatsinst. Exp. Ther. Frankfurt a. M.*, **22**, 31 (1929); *Ibid.*, **23**, 65–80 (1930).

INBREEDING AND HOMOZYGOSIS

BY SEWALL WRIGHT

DEPARTMENT OF ZOÖLOGY, THE UNIVERSITY OF CHICAGO

Communicated February 23, 1933

The elementary evolutionary process in a Mendelian population is change of gene frequency. The pressures of immigration, selection and mutation on gene frequency are supplemented by the random drift brought about by the accidents of sampling among the germ cells of each generation in any population of finite size. This last process is important as furnishing a basis for the nonadaptive differentiation of local races.

The rate at which the process takes place is measured by the rate of decrease of heterozygosis for paired allelomorphs, not subject to other evolutionary pressure. This rate has been calculated for systems of close inbreeding by working out the consequences of every possible type of mating and giving each type its due weight in each generation. The cases investigated in this way include self-fertilization,[1] parent-offspring mating,[2] for both autosomal and sex-linked genes, and brother-sister mating,[2] also for both autosomal and sex-linked genes. This method is not practicable for larger inbreeding groups. Another method, that of path coefficients has been applied to a variety of systems of mating in the case of autosomal genes.[3,4,5] It is the purpose of the present paper to extend these results to the case of sex-linked genes.

It may be desirable to review the method of path coefficients briefly. In general the object of investigation is a system of variable quantities which for any reason it is appropriate to treat as occupying positions in a network of relations of cause and effect. Such a system is conveniently represented in a diagram like figure 1. Those variables which are treated as dependent are connected with those of which they are considered as functions by arrows. The system of factors back of each variable may be made formally complete by the introduction of symbols representative of total residual determination. A residual correlation between variables is represented by a double headed arrow. It is assumed that all relations are linear. Thus each variable is related to those from which unidirectional arrows are drawn to it by an equation of the following type in which $(V_0 - \overline{V}_0)$, $(V_1 - \overline{V}_1)$, etc., represent deviations from the means and c_{01}, c_{02} are the coefficients.

$$(V_0 - \bar{V}_0) = c_{01}(V_1 - \bar{V}_1) + c_{02}(V_2 - \bar{V}_2) \ldots + c_{0n}(V_n - \bar{V}_n) \quad (1)$$

It is convenient to measure the deviation of each variable by its standard deviation. Let

$$X_0 = \frac{V_0 - \bar{V}_0}{\sigma_0}, \quad X_1 = \frac{V_1 - \bar{V}_1}{\sigma_1}, \quad \text{etc.}$$

and let

$$p_{01} = c_{01}\frac{\sigma_1}{\sigma_0}, \quad p_{02} = c_{02}\frac{\sigma_2}{\sigma_0}, \quad \text{etc.}$$

Then

$$X_0 = p_{01}X_1 + p_{02}X_2 \ldots + p_{0n}X_n. \quad (2)$$

The coefficients in this form are of the type called path coefficients. The theorem which makes them of value in relating correlations to causal relations is a very simple one. The correlation between V_0 and any other variable, such as V_q in figure 2 can be written in the form:

$$\begin{aligned} r_{0q} &= \frac{\Sigma X_0 X_q}{N} = \frac{1}{N}\Sigma X_q(p_{01}X_1 + p_{02}X_2 \ldots p_{0n}X_n) \\ &= p_{01}r_{qi} + p_{02}r_{q2} \ldots p_{0n}r_{qn} \\ &= \Sigma p_{0i}r_{qi}. \end{aligned} \quad (3)$$

The correlation is thus analyzed into contributions from each of the paths in the diagram (Fig. 2) by which the two variables are connected through common factors.

But the correlation terms of type r_{qi} may be capable of analysis by application of the same formula. By repeated analysis of this sort as far as

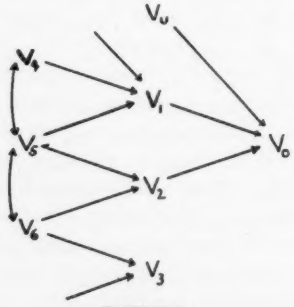


FIGURE 1

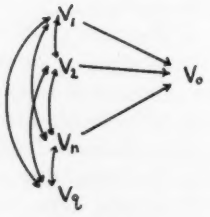


FIGURE 2

the diagram (such as Fig. 1) permits, we are led to the following principle. Any correlation between variables in a network of sequential relations can be analyzed into contributions from each of the paths (direct or through a common *factor*) by which the two variables are connected, such

that the value of each contribution is the product of the coefficients pertaining to the elementary paths. Only one of the coefficients involved in each of these products can be a correlation coefficient. The rest are path coefficients. In tracing connecting paths, it is obvious that one may trace back along the arrows and then forward, as well as forward from one variable to the other, but never forward and then back. The double headed arrows imply unknown common factors. As illustrations, consider the correlations between some of the variables in figure 1.

$r_{46} = 0$ (no connecting path shown)
$r_{36} = p_{36}$ (only one connecting path)
$r_{13} = p_{15}r_{56}p_{36}$ (one connecting path)
$r_{23} = p_{25}r_{56}p_{36} + p_{26}p_{36}$ (two connecting paths)
$r_{12} = p_{14}r_{45}p_{25} + p_{15}p_{25} + p_{15}r_{56}p_{26}$ (three connecting paths).

An important special case of equation 3 arises if one correlates a variable with itself taking account of *all* factors (known and unknown).

$$r_{00} = \Sigma p_{0i}r_{0i} = 1. \tag{4}$$

If merely known factors are used, $\Sigma p_{0i}r_{0i} = r^2_{0(12\ldots i,\ldots n)}$ the squared multiple correlation coefficient. Equation 4 may be put in a form which is often more convenient by further analysis of the correlation term.

$$\Sigma p^2_{0i} + 2\Sigma p_{0i}p_{0j}r_{ij} = 1. \tag{5}$$

The squared path coefficients measure the portion of the variance of the dependent variable for which the factor in question is directly responsible under the point of view adopted. The terms of the type $2p_{0i}p_{0j}r_{ij}$ give the effect (positive or negative) on the variance, due to correlation between the indicated factors.

The Fixation of Sex-Linked Genes by Inbreeding

In figure 3, individuals of the homogametic sex are represented by circles and will be referred to as females (as in mammals and flies, but not birds or moths) while individuals of the heterogametic sex are represented by squares and will be referred to as males. With respect to sex-linked genes, the constitution of a male is completely determined by that of the unfertilized egg and completely determines that of any sperm which he produces. The corresponding path coefficients accordingly have the value 1. The other path coefficients and the correlation coefficients are represented by letters in figure 3. Primes are used to indicate generations preceding that in question. Correlations which are analyzed in the figure (f, m, f') are indicated by dotted arrows (two headed).

The frequency of sex-linked genes is not necessarily the same in males and females but there is rapid oscillatory approach to identity on random

mating.[2] Assuming identity, the constitution of a female is equally and completely determined by the genes of the egg and sperm which united to produce it. Thus by equation (5)

$$2a^2 + 2a^2f = 1$$

$$a = \sqrt{\frac{1}{2(1+f)}}. \qquad (6)$$

The individuals of a given generation are also equally and completely determined (in a mathematical sense) by pairs of complementary germ cells which they produce (egg and polar body from the reduction division). Thus the correlation between individual constitution and germ cell produced (b) must be the same (in the absence of selection or other disturbing factor) as that between individual and one of the germ cells from which it came ($a' + a'f'$).

$$b = \sqrt{\frac{1}{2(1+f')}}\,(1+f') = \sqrt{{}^1/_2(1+f')}. \qquad (7)$$

It may be seen that ba' measuring the influence of germ cell of one generation on one of the next, has the value $^1/_2$ irrespective of inbreeding.

The correlation between uniting gametes (f) is closely related to the percentage of heterozygosis. Assume that genes A and a are present in the frequencies q and $(1-q)$ in the total population composed of many inbred lines and that the total percentage of heterozygosis is p. By the product moment formula for correlation

	A	a	Total
A	$q - p/2$	$p/2$	q
a	$p/2$	$1 - q - p/2$	$1-q$
Total	q	$1-q$	1

$$f = \frac{(q - p/2) - q^2}{q(1-q)} = 1 - \frac{p}{2q(1-q)}$$

$$p = 2q(1-q)(1-f).$$

Since the proportion of heterozygosis in a random bred stock (total population before subdivision into inbred lines) is $p_0 = 2q(1-q)$

$$p = p_0(1-f). \qquad (8)$$

From inspection of figure 3, we have under brother-sister mating

$$f = ba'(f' + b'^2)$$

$$= {}^1/_4(1 + 2f' + f'') \qquad (9)$$

The value of f can be written for any number of generations by this formula. Substituting the value of p from equation 8,

$$p = {}^1/_2\,p' + {}^1/_4\,p''. \qquad (10)$$

These formulae agree with the results which Jennings[2] obtained for this case by working out the consequences of every possible mating. They happen to be identical with those applying to autosomal genes under the same system of mating. The mean rate of decrease of heterozygosis (putting $p/p' = p'/p''$) is $1/4(3 - \sqrt{5}) = 19.1\%$.

It may be noted that there is nothing in the reasoning which limits the number of individuals of the heterogametic sex to one in each line in each generation. In flocks of poultry headed always by one male (homogametic in birds) but with an indefinite number of females, the fixation of sex-linked genes proceeds as rapidly as under full brother-sister mating although autosomal genes are fixed only about half as rapidly under this system of mating (11.0% per generation).

We will now turn to the more general case of populations of N_m males and of N_f females. Two types of matings are to be considered, those in which the mated individuals have the same mother (Fig. 3) and those in which they have different mothers (Fig. 4). The frequency of the former is $\frac{1}{N_f}$, of the latter $(1 - 1/N_f)$.

Case (1), figure 3. $f_{(1)} = 1/4(1 + 2f' + f'')$

Case (2), figure 4. $f_{(2)} = 1/2(f' + f_f')$

where f_f' is the correlation between eggs produced by different females of the same generation.

Combining
$$f = \frac{1}{N_f} f_{(1)} + (1 - 1/N_f) f_{(2)}$$
$$= 1/2(f' + f_f') + 1/4 N_f(1 - 2f_f' + f''). \qquad (11)$$

For a solution it is necessary to determine f_f. It will be found that this involves the correlation between sperms of different males. It will be convenient to consider this first.

Again there are two cases to be considered.

Case (1), figure 5 (mother in common), frequency $1/N_f$.

$$f_{m(1)} = b'^2 = 1/2(1 + f'').$$

Case (2), figure 6 (mother not in common), frequency $1 - 1/N_f$

$$f_{m(2)} = f_f'.$$

Total
$$f_m = \frac{1}{N_f} f_{m(1)} + (1 - 1/N_f) f_{m(2)}$$
$$= f_f' + \frac{1}{2N_f}(1 - 2f_f' + f'')$$
$$= 2f - f' \text{ (from equation 11).} \qquad (12)$$

There are four cases to be considered in the case of f_f.

Case (1), full brother and sister (Fig. 7). Frequency $\frac{1}{N_m N_f}$

$$f_{f(1)} = b^2 a'^2(1 + b'^2 + 2f')$$
$$= {}^1/_8(3 + 4f' + f'').$$

Case (2), mother only in common (Fig. 8). Frequency $\frac{N_m - 1}{N_m N_f}$

$$f_{f(2)} = b^2 a'^2(b'^2 + 2f' + f'_m)$$
$$= {}^1/_8(1 + 8f' - f'').$$

Case (3), father only in common (Fig. 9). Frequency $\frac{N_f - 1}{N_m N_f}$

$$f_{f(3)} = {}^1/_4(1 + 2f' + f'_f)$$

Case (4), neither parent in common (Fig. 10). Frequency $\frac{(N_m - 1)(N_f - 1)}{N_m N_f}$

$$f_{f(4)} = {}^1/_4(f'_m + f'_f + 2f')$$
$$= {}^1/_4(4f' - f'' + f'_f).$$

$$\text{Total } f_f = \frac{1}{N_m N_f} f_{f(1)} + \frac{(N_m - 1)}{N_m N_f} f_{f(2)} + \frac{(N_f - 1)}{N_m N_f} f_{f(3)} + \frac{(N_m - 1)(N_f - 1)}{N_m N_f} f_{f(4)}$$
$$= {}^1/_4(2f + 3f' - f'') + {}^1/_4 N_m(1 - 2f' + f''). \quad (13)$$

Substituting in equation (11) we reach after some reduction

$$f = f' + \frac{(N_f + 1)}{8N_f}(1 - 2f' + f'') - \frac{(N_m - 1)(N_f - 1)}{8N_m N_f}(1 - 2f'' + f'''). \quad (14)$$

$$p = p' - \frac{(N_f + 1)}{8N_f}(2p' - p'') + \frac{(N_m - 1)(N_f - 1)}{8N_m N_f}(2p'' - p'''). \quad (15)$$

These are the required iteration formulae for f and p. It may be seen that if there is only one individual of the homogametic sex per generation in each inbred group ($N_f = 1$), equations (14) and (15) reduce to (9) and (10) as required. If there is only one individual of the heterogametic sex ($N_m = 1$) while N_f is large, $p = {}^3/_4 p' + {}^1/_8 p''$, which is the same as for autosomal genes in herds headed always by one male.

The average rate of decrease of heterozygosis can be obtained by equating p/p', p'/p'' and p''/p''' and expressing (15) in terms of $x = (p/p') - 1$

$$x^3 + x^2(2 + 2c_1) + x(1 + 3c_1 - 2c_2) + (c_1 - c_2) = 0 \quad (16)$$

$$\text{where } c_1 = \frac{N_f + 1}{8N_f} \quad \text{and} \quad c_2 = \frac{(N_m - 1)(N_f - 1)}{8N_mN_f}.$$

As noted above, if either $N_f = 1$ or $N_m = 1$, $c_2 = 0$ and formula 16 reduces to $x^2 + x(1 + 2c_1) + c_1 = 0$ of which the solution is

$$x = \frac{\Delta p}{p'} = -\frac{1}{2}[1 + 2c_1 - \sqrt{1 + 4c_1^2}]. \qquad (17)$$

With one female this gives $-1/4(3 - \sqrt{5}) = -19.1\%$.
With one male it varies from -19.1% (brothers with sisters) to

FIGURE 3

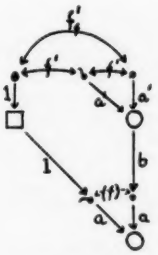


FIGURE 4

FIGURE 5

FIGURE 6

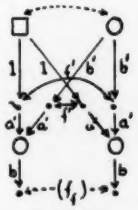


FIGURE 7

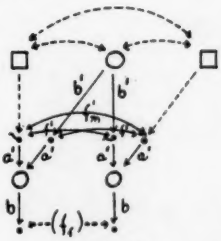


FIGURE 8

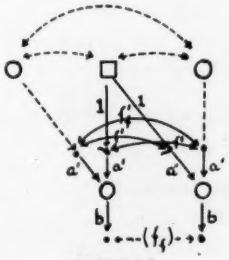


FIGURE 9

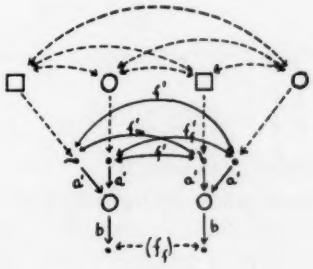


FIGURE 10

$-\frac{1}{8}(5-\sqrt{17}) = -11.0\%$ (one male with an indefinitely large number of half sisters).

For reasonably large values of N_m and N_f, x becomes so small that the powers higher than the first in (16) may be ignored leading to the approximate solution

$$\frac{\Delta p}{p'} = -\left(\frac{2N_m + N_f}{9N_mN_f}\right).$$

These and previous results are brought together for comparison in the following list.

I. Random mating in dioecious populations of limited size.[5]

A. Autosomal genes	p	$-\frac{\Delta p}{p'}$ (approximation)
N_m ♂'s, N_f ♀'s	$p' - \left(\frac{N_m + N_f}{8N_mN_f}\right)(2p' - p'')$	$\frac{1}{8N_m} + \frac{1}{8N_f}$
N_m ♂'s, ∞ ♀'s	$p' - \frac{1}{8N_m}(2p' - p'')$	$\frac{1}{8N_m + 1}$
$N/2$ ♂'s, $N/2$ ♀'s	$p' - \frac{1}{2N}(2p' - p'')$	$\frac{1}{2N+1}$
1 ♂, ∞ ♀'s	$\frac{3}{4}p' + \frac{1}{8}p''$	11.0%
1 ♂, 1 ♀	$\frac{1}{2}p' + \frac{1}{4}p''$	19.1%
B. Sex-linked genes.		
N_m ♂'s, N_f ♀'s	$p' - \left(\frac{N_f + 1}{8N_f}\right)(2p' - p'') + \frac{(N_m - 1)(N_f - 1)}{8N_mN_f}(2p'' - p''')$	$\frac{2}{9N_f} + \frac{1}{9N_m}$
$N/2$ ♂'s, $N/2$ ♀'s	$p' - \left(\frac{N+2}{4N}\right)(2p' - p'') + \frac{(N-2)^2}{8N^2}(2p'' - p''')$	$\frac{2}{3N}$
1 ♂, ∞ ♀'s	$\frac{3}{4}p' + \frac{1}{8}p''$	11.0%
N_m ♂'s, 1 ♀ / 1 ♂, 1 ♀	$\frac{1}{2}p' + \frac{1}{4}p''$	19.1%

II. Monoecious populations[5]

(N individuals with random union of gametes), $\frac{2N-1}{2N}p'$ $\frac{1}{2N}$ exactly

$N = 1$ (self-fertilization), $^1/_2 p'$	50% exactly
No self-fertilization, $p' - \frac{1}{2N}(2p' - p'')$	$\frac{1}{2N+1}$

III. Limited populations with maximum possible avoidance of consanguineous mating.[3]

$N = 4$ (double first cousins), $p' - (^1/_{16})p^{IV}$	8.0%
$N = 8$ (quadruple second cousin), $p' - (^1/_{32})p^{V}$	3.5%
$N = 16$ (octuple third cousin), $p' - (^1/_{64})p^{VI}$	1.7%
$N = 2^m$ (large), $p' - \left(\frac{1}{4N}\right)p^{(m+2 \text{ primes})}$ (by analogy)	$\frac{1}{4N}$

IV. Parent and offspring[2,3]

Grading up to homozygote (autosomal), $p'/2$	50%
Grading up to heterogametic sex (sex-linked), $p'/2 + p''/4$	19.1%
Father-daughter, mother-son (autosome), $p'/2 + p''/4$	19.1%
Father-daughter, mother-son (sex-linked), $p''/2$	29.3%

V. Unlimited population with avoidance of inbreeding except of specified sort[3]

Half brother-sister $p_\infty = 0$, approach at continually decreasing percentage rate, $p_5 = 0.68\,p_0$, $p_{10} = 0.56\,p_0$, $p_{15} = 0.49\,p_0$

First cousins $p_\infty = 0$, approach at decreasing percentage rate, $p_5 = 0.83\,p_0$, $p_{10} = 0.75\,p_0$, $p_{15} = 0.69\,p_0$

Half first cousins $p = {}^1/_{32}(26\,p_0 + 4p'' + p'''')$, $p_\infty = 26/27\,p_0 = 0.96\,p_0$

Second cousins $p = {}^1/_{64}(52\,p_0 + 8\,p'' + 2\,p'''' + p^{IV})$, $p_\infty = 52/53\,p_0 = 0.98\,p_0$

VI. Assortative mating[3] in character dependent on n duplicate genes lacking dominance.

Limit, $p_\infty = \frac{n(1-m)}{2n(1-m)+m}$ when $m = h^2_\infty r$, m being genetic assortative mating and h^2_∞ the limiting degree of determination of variance by heredity.

VII. Irregular pedigrees[4] (live stock).

In general, $f = \Sigma\,[(^1/_2)^{n_s+n_d+1}(1+f_A)]$, $p = p_0(1-f)$

where n_s and n_d are the number of generations from the sire and dam, respectively to a common tie (A). This formula lends itself to sampling methods[6] making possible a study of the histories of whole breeds.[7,8,9,10]

A very rapid decrease in heterozygosis is naturally indicated in the formative period of the breed (studied in Shorthorn cattle[7,8]). But even after large numbers are reached the average rate of decrease per generation in 6 breeds including cattle,[8,10] horses,[9] sheep[10] and swine[10] is about 0.5%, equivalent to 25 as the effective number of sires per generation. It is probably also true in nature that the effective population number may be very much smaller than the actual number, making the effect considered here an important one in the life of the species.

[1] Jennings, H. S., *Amer. Nat.*, **45**, 487–497 (1912).
[2] Jennings, H. S., *Genetics*, **1**, 53–89 (1916).
[3] Wright, S., *Ibid.*, **6**, 111–178 (1921).
[4] Wright, S., *Amer. Nat.*, **61**, 330–338 (1922).
[5] Wright, S., *Genetics*, **16**, 97–159 (1931).
[6] Wright, S., and McPhee, H. C., *J. Agr. Res.*, 377–383 (1925).
[7] Wright, S., *J. Hered.*, **14**, 339–349, 405–422 (1923).
[8] McPhee, H. C., and Wright, S., *Ibid.*, **16**, 205–215 (1925); *Ibid.*, **17**, 397–401 (1926).
[9] Calder, A., *Proc. Roy. Soc. Edinb.*, **47**, 118–140 (1927).
[10] Lush, J. L., *Proc. 6th Internat. Cong. Genetics*, **2**, 123–126 (1932).

INBREEDING AND RECOMBINATION

By Sewall Wright

Department of Zoölogy, The University of Chicago

Communicated February 23, 1933

It is often asserted that a correlation between two characters in a random breeding population may be due either to a common physiological factor or to genetic linkage. The latter possibility was investigated by Robbins[1] (1918) for a population of mixed origin. He found that there is a rather rapid approach to the proportions of random combination, unless linkage is very strong. Equilibrium is not, however, reached immediately, even without linkage, as pointed out long ago by Weinberg.[2]

The effect of inbreeding on such recombination was worked out by Robbins[1] in the simplest case, that of self-fertilization. Recently Haldane and Waddington[3] have studied the effects of systems of inbreeding, involving two individuals per generation (parent-offspring and brother-sister, for autosomal and sex-linked genes in each case). They used a somewhat shorter method than Robbins, but one that still depends on the working out of the consequences of every possible type of mating. Brother-sister mating involved consideration of 100 different types of mating, which could be grouped in 22 classes. With regard to inbreeding within larger groups they say:

"The method of calculation here employed cannot be applied to most other types of inbreeding. It is wholly inapplicable to the important case where a male is bred to a large group of his half-sisters in each generation. The case of double first cousin matings with autosomal linkage involves the consideration of 10,000 different pairs of mating types, and other systems are still more complex. It may prove possible to solve such problems by an extension of Wright's[4] (1921) correlation method, but we have been unable to do so."

The purpose of the present paper is to consider the applicability of the correlation method referred to above, to this problem. The method (that of path coefficients) has been reviewed briefly in a preceding paper.[5]

We will begin with the effects of random mating, in an indefinitely large population, on two pairs of autosomal genes.

s = the correlation between the two pairs of genes in two germ cells of the same individual.

r_m, r_f = the correlation between the two pairs of genes in the *same* sperm cell (r_m) or egg cell (r_f).

c_m, c_f = the recombination frequencies.

a = path coefficient measuring the influence of gamete on fertilized egg ($= \sqrt{1/2}$ under random mating).

b = path coefficient measuring influence of zygote on gametes produced by it ($= \sqrt{1/2}$ under random mating).

Primes are used to designate preceding generations. It has been shown that $ba' = 1/2$ irrespective of inbreeding.

The accompanying table is intended to illustrate the nature of the correlation coefficients, r (and s). The combinations AB, Ab, aB and ab are assumed to occur in the frequencies w, x, y and z, respectively (whether the A series and B series are in the same gamete (r) or in different gametes (s)). The gene arrays are $(1 - q_A)a + q_A A$ and $(1 - q_B)b + q_B B$, respectively. By the product moment formula for correlation

	B	b	Total
A	w	x	q_A
a	y	z	$1 - q_A$
Total	q_B	$1 - q_B$	1

$$r = \frac{w - q_A q_B}{\sqrt{q_A q_B (1 - q_A)(1 - q_B)}} = \frac{wz - xy}{\sqrt{q_A q_B (1 - q_A)(1 - q_B)}} \qquad (1)$$

Thus $r = 0$ if $w = q_A q_B$, the proportion of random combination.

Tracing the connecting paths in figure 1

$$s = b^2 a'^2 (r_m' + r_f') = \frac{1}{2} F'$$

With random assortment ($c = 1/2$), $r_m = r_f = s = \frac{1}{2} F'$. In this case the departure from random combination of genes is halved in each successive generation as found by Weinberg and by Robbins.

With linkage, there will be a tendency for the two series of genes to remain in the same association as that in which they entered the individual. This tendency may be different in males and females depending on the

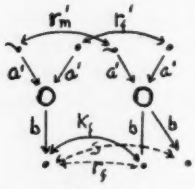


FIGURE 1

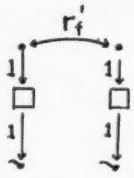


FIGURE 2

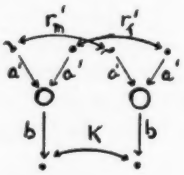


FIGURE 3

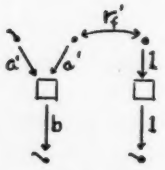


FIGURE 4

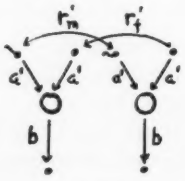


FIGURE 5

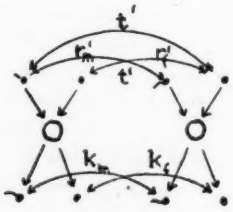


FIGURE 6

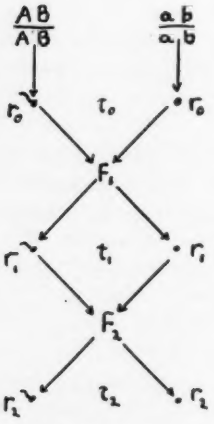


FIGURE 7

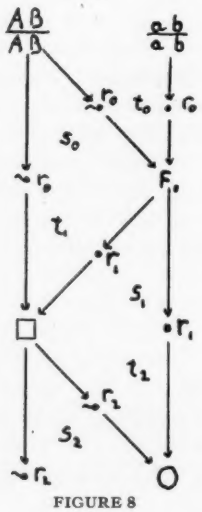


FIGURE 8

recombination values c_m and c_f and is measured by k_m and k_f in the two cases

$$r_m = s + k_m$$
$$r_f = s + k_f$$

If $c_m = 0$, $r_m = \bar{r}'$ and if $c_f = 0$, $r_f = \bar{r}'$. For values of c_m between 0 and $1/2$, r_m must take values interpolated linearly between $\bar{r}'$ and $\frac{1}{2}\bar{r}'$.

$$\left.\begin{aligned} r_m &= \tfrac{1}{2}\bar{r}' + (1/2 - c_m)\bar{r}' = (1 - c_m)\bar{r}' \\ r_f &= \tfrac{1}{2}\bar{r}' + (1/2 - c_f)\bar{r}' = (1 - c_f)\bar{r}' \\ \bar{r} &= (1 - \bar{c})\bar{r}' \end{aligned}\right\} \quad (2)$$

$$\frac{\Delta\bar{r}}{\bar{r}'} = \frac{\bar{r} - \bar{r}'}{\bar{r}'} = -\bar{c} \quad (3)$$

Thus $\bar{r}$ decreases at the rate of the mean recombination frequency ($100\,\bar{c}\,\%$) per generation as it approaches its limiting value, $r_\infty = 0$, in agreement with Robbins' results. With respect to two *characters* (A, B) whose variations are determined completely ($h^2 = 1$) and linearly (no dominance) by the two pairs of genes, respectively,

$$r_{AB} = a^2(r_m + r_f) = \bar{r} \quad (4)$$

If determination is incomplete, $r_{AB} = h_A h_B \bar{r}$. Dominance also reduces the correlation. Thus the correlation of characters due to racial mixture falls off at the rate $100\,\bar{c}\,\%$ per generation and ultimately disappears.

The situation is a little more complicated in the case of two sex-linked genes. Treating the male as the heterogametic sex, the system of relations are as represented in figures 2 and 3 (males represented by squares, females by circles).

The constitution of the single X-chromosome of a male is completely determined in the unfertilized egg derived from the mother and completely determines the constitution of the X-chromosome of the sperms produced. Thus $b_m = a'_m = 1$ in this case and $r_m = r'_f$. The X-chromosomes of females are derived equally from the parents and $b_f a'_f = 1/2$ as with autosomal genes.

It will be convenient to drop the subscripts for the coefficients pertaining to the females, writing r for r_f, etc.

$$r = 1/4(r' + r'') + k$$

If $c = 0$ $\quad r = 1/2(r' + r''),\ k = 1/4(r' + r'')$.

If $c = 1/2,\ k = 0$ $\quad r = 1/4(r' + r'')$

Interpolating $\quad r = (r' + r'')\,\dfrac{(1 - c)}{2}$ (5)

This is the required iteration formula for r. The mean percentage rate of change can be found by equating $\frac{r}{r'}$ to $\frac{r'}{r''}$ and solving for $\frac{\Delta r}{r'} = \left(\frac{r}{r'}\right) - 1$.

$$\left(\frac{\Delta r}{r'}\right)^2 + \left(\frac{\Delta r}{r'}\right)\left(\frac{3+c}{2}\right) + c = 0$$

$$\left(\frac{\Delta r}{r'}\right) = -\frac{1}{12}\left[8c + (9-5c)\left(1 - \sqrt{1 - \left(\frac{4c}{9-5c}\right)^2}\right)\right] \qquad (6)$$

Approximately $\frac{\Delta r}{r'} = -\frac{2c}{3}\left[1 + \frac{c}{9-5c}\right]$. This is to be compared to the rate $\frac{\Delta r}{r'} = -c$ for autosomal genes.

The next case is that of two genes, one of which is autosomal and the other sex-linked. (Figs. 4, 5.) Writing r for r_f,

$$r_m = \frac{1}{2} r_f' = \frac{1}{2} r'$$

$$r = 1/4(r' + r'_m) = 1/8(2r' + r'') \qquad (7)$$

This is the required iteration formula. The rate of change of r comes out $\frac{\Delta r}{r'} = -1/2$. This is the same as for two unlinked autosomal genes.

Self-fertilization. (Fig. 6.)

$t = 1/2(\bar{r}' + t')$ ($= r_{AB}$ in uniting gametes)
$r_m = 1/2(\bar{r}' + t') + k_m$ ($= r_{AB}$ in same sperm)
$r_f = 1/2(\bar{r}' + t') + k_f$ ($= r_{AB}$ in same egg)
If $c_m = 0$, $r_m = \bar{r}'$, but if $c_m = {}^1/_2$, $r_m = {}^1/_2(\bar{r}' + t')$

Interpolating

$$r_m = \bar{r}' - c_m\bar{k}', \quad k_m = \left(\frac{1-2c_m}{2}\right)\bar{k}'$$

$$r_f = \bar{r}' - c_f\bar{k}', \quad k_f = \left(\frac{1-2c_f}{2}\right)\bar{k}'$$

$$\bar{r} = \bar{r}' - \bar{c}\,\bar{k}', \quad \bar{k} = \left(\frac{1-2\bar{c}}{2}\right)\bar{k}'$$

$$\sum_2^n \bar{r} - \sum_1^{n-1} \bar{r} = -\bar{c}\sum_1^{n-1} \bar{k}$$

$$\bar{r}_n - \bar{r}_1 = -\bar{c}\bar{k}_1\left[\frac{1 - \left(\frac{1-2\bar{c}}{2}\right)^{n-1}}{1 - \left(\frac{1-2\bar{c}}{2}\right)}\right]$$

$$r_\infty = \bar{r}_1 - \frac{2\bar{c}\,\bar{k}_1}{1+2\bar{c}} \qquad (8)$$

Assume that the initial population is the result of a cross between strains differing in two genes ($AABB \times aabb$) and that r_0 and t_0 (Fig. 7) refer to gametes from these strains. Obviously $r_0 = 1$ while $t_0 = -1$ (as the correlation between A and B in *uniting* gametes). The values for the gametes of the cross bred individuals must also be obtained from consideration of the diagram, rather than by blind application of the formulae which apply after self-fertilization has begun. Thus $t_1 = {}^1/_4(2\bar{r}_0 + 2t_0) = 0$; $\bar{r}_1 = 1 - 2\bar{c}$ (by interpolation between the values 0 and 1 which obviously hold when $\bar{c} = {}^1/_2$ and 0, respectively; $\bar{k}_1 = \bar{r}_1 - t_1 = 1 - 2\bar{c}$.

The most important conclusion to be drawn is the value of r, reached when complete homozygosis has been obtained in all branches of the original population. By substitution:

$$r_\infty = \frac{1 - 2\bar{c}}{1 + 2\bar{c}} \tag{9}$$

Closely related is the proportion of recombination classes in this limiting population $(x + y)$. In the present case in which $q_A = q_B = {}^1/_2$, $x + y = 2x = 1 - 2w$. Since

$$\begin{aligned} r &= 4w - 1 = 1 - 4x \\ 2x &= \frac{1 - r}{2} \qquad (10) \\ 2x_\infty &= \frac{2\bar{c}}{1 + 2\bar{c}} \end{aligned}$$

These results are identical with those obtained by Robbins and by Haldane and Waddington.

In the following cases we will use c for the mean recombination proportion.

Parent and offspring, autosomal genes

$$\begin{aligned} s &= {}^1/_4(r' + r'' + 2t') \quad (= r_{AB} \text{ for gametes of same individual}) \\ t &= {}^1/_2(s' + t') \quad (= r_{AB} \text{ for uniting gametes}) \\ r &= s + k \quad (= r_{AB} \text{ in same gametes}) \\ &= {}^1/_2(r' + r'') - \frac{c}{2}(r' + r'' - 2t') \text{ (interpolation method)} \\ k &= \left(\frac{1 - 2c}{4}\right)(r' + r'' - 2t') = (1 - 2c)(s - t') \\ &= \left(\frac{1 - 2c}{4}\right)(k' + k'' + s' - t'') \\ &= \left(\frac{1 - 2c}{4}\right)\left(k' + k'' + \frac{k'}{1 - 2c}\right) \end{aligned}$$

$$= k'\left(\frac{1-c}{2}\right) + k''\left(\frac{1-2c}{4}\right)$$

$$r = {}^1/_2(r' + r'') - \frac{2ck}{1-2c}$$

$$2\sum_3^n r - \sum_2^{n-1} r - \sum_1^{n-2} r = -\frac{4c}{1-2c}\sum_3^n k$$

But $$\sum_3^n r = r_n + \sum_2^{n-1} r - r_2 = r_n + r_{n-1} + \sum_1^{n-2} r - r_1 - r_2$$

$$3r_\infty - 2r_2 - r_1 = -\frac{4ck}{1-2c}\sum_3^\infty k$$

$$\sum_3^\infty k = \sum_2^\infty k\left(\frac{1-c}{2}\right) + \sum_1^\infty k\left(\frac{1-2c}{4}\right)$$

$$= k_1\left(\frac{1-2c}{1+4c}\right) + k_2\left(\frac{3-4c}{1+4c}\right)$$

Finally $$3r_\infty = 2r_2 + r_1 - \frac{4c}{1+4c}\left[k_1 + k_2\left(\frac{3-4c}{1-2c}\right)\right] \qquad (11)$$

Case 1.—Initial population from $AABB \times aabb$ followed by a mating of F_1 to parents, etc. (Fig. 8.)

$$s_0 = 1 \quad s_1 = 0 \quad s_2 = {}^1/_2(1-c)$$
$$t_0 = -1 \quad t_1 = 0 \quad t_2 = 0$$
$$r_0 = 1 \quad r_1 = 1-2c \quad r_2 = (1-c)^2$$
$$k_0 = 0 \quad k_1 = 1-2c \quad k_2 = {}^1/_2(1-c)(1-2c)$$

$$r_\infty = \frac{3-4c}{3(1+4c)}$$

$$2x_\infty = \frac{8c}{3(1+4c)} \qquad (12)$$

Haldane and Waddington give $r_\infty = \frac{1}{1+4c}$, $2x_\infty = \frac{2c}{1+4c}$ for this case.

Case 2.—$AABB \times aabb$, $F_1 \times F_1$, $F_2 \times F_1$, etc. (Fig. 9.)

$$s_0 = 1 \quad s_1 = 0 \quad s_2 = {}^1/_2(1-2c) \quad s_3 = {}^1/_4(1-2c)(2-c)$$
$$t_0 = -1 \quad t_1 = 0 \quad t_2 = 0 \quad t_3 = {}^1/_4(1-2c)$$
$$r_0 = 1 \quad r_1 = 1-2c \quad r_2 = (1-2c)(1-c) \quad r_3 = {}^1/_2(1-2c)(2-c)^2$$
$$k_0 = 0 \quad k_1 = 1-2c \quad k_2 = {}^1/_2(1-2c)^2 \quad k_3 = {}^1/_4(1-2c)^2(2-c)$$

$$r_\infty = \frac{1-2c}{1+4c} \qquad 2x_\infty = \frac{3c}{1+4c} \qquad (13)$$

These agree with the results of Haldane and Waddington.

Parent and offspring, sex-linked genes

The diagram is essentially the same as that for self-fertilization, if one considers only the female line. Heterozygosis is halved from mothers to daughters. Starting from double heterozygotes mated to their sons (from

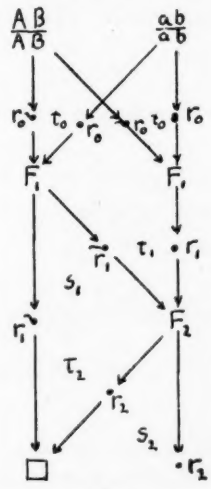


FIGURE 9

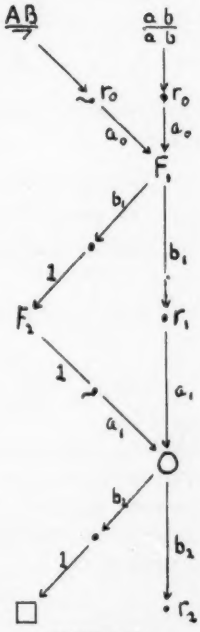


FIGURE 10

any fathers) the relations of two series of sex-linked genes is the same as under self-fertilization. (Fig. 10.)

$$r_\infty = \frac{1 - 2\bar{c}}{1 + 2\bar{c}}, \qquad 2x_\infty = \frac{2\bar{c}}{1 + 2\bar{c}} \tag{14}$$

Haldane and Waddington give

$$r_\infty = \frac{6 - 7\bar{c}}{6 + 9\bar{c}}, \qquad 2x_\infty = \frac{8\bar{c}}{6 + 9\bar{c}}$$

These are clearly in error since for $c = {}^1/_2$, the eggs, and hence also the sons, of heterozygous females must show random assortment ($r_\infty = 0$, $2x_\infty = {}^1/_2$) while their formulae yield $r_\infty = \frac{5}{21}$, $2x_\infty = \frac{8}{21}$ in this case.

N_m males, N_f females—autosomal genes

Random mating is assumed within the group. The correlation (t) between A and B of two random germ cells is the same in a given generation whether two eggs, two sperms, or egg and sperm are compared, there being no difference between males and females in the frequencies of the various combinations. This is also true of the correlation (s) between A and B in two germ cells of the same individual, whether male or female.

$$s = {}^1/_4(r'_m + r'_f + 2t') = {}^1/_2(\bar{r}' + t')$$

Three cases must be considered with respect to germ cells from different individuals.

Case (1)	Both parents in common	$t_{(1)} = {}^1/_4(2s' + 2t')$
Case (2)	One parent in common	$t_{(2)} = {}^1/_4(s' + 3t')$
Case (3)	Neither parent in common	$t_{(3)} = {}^1/_4(4t')$

$$\text{Total} \quad t = \frac{1}{4N_mN_f}\, t_{(1)} + \frac{N_m + N_f - 2}{N_mN_f}\, t_{(2)} + \frac{(N_m - 1)(N_f - 1)}{N_mN_f}\, t_{(3)}$$

$$t = t' + \left(\frac{N_m + N_f}{4N_mN_f}\right)(s' - t')$$

Differences in crossing-over in spermatogenesis and oögenesis make a difference in the correlation (r) between A and B in the same sperm from that in the same egg, but as before the average value of r, k and c can be used in the analysis, the different values for r_m and r_f, etc., being derivable at any desired point.

$$r = s + k$$

If $c = 0$, $r = r'$ and if $c = {}^1/_2$, $r = {}^1/_2(r' + t')$ giving by linear interpolation

$$r = r' - c(r' - t')$$

and

$$k = \frac{(1 - 2c)}{2}(r' - t')$$

$$r' - t' = \frac{2k}{1 - 2c}$$

Also

$$r' - t' = k' + (s' - t')$$

But

$$s' - t' = {}^1/_2(r'' - t'') - \sigma(s'' - t'') \quad \text{where } \sigma = \frac{N_m + N_f}{4N_mN_f}$$

$$= \frac{k'}{1 - 2c} - \sigma(s'' - t'')$$

Thus $$\frac{2k}{1-2c} = k' + \frac{k'}{1-2c} - \sigma(s'' - t'')$$

$$\frac{2\sigma k'}{1-2c} = \sigma k'' + \sigma(s'' - t'')$$

Adding $\left(\frac{2}{1-2c}\right)(k + \sigma k') = k' + \frac{k'}{1-2c} + \sigma k''$

$$k = k'(1 - c - \sigma) + k''\sigma\left(\frac{1-2c}{2}\right)$$ iteration formula for k

$$r - r' = -\frac{2ck}{1-2c}$$

$$\sum_1^n r - \sum_0^{n-1} r = -\frac{2c}{1-2c}\sum_1^n k$$

$$r_\infty = 1 - \frac{4c}{1-2c}\left[\frac{k_1(c+\sigma) + k_2}{2c + 2c\sigma + \sigma}\right] \qquad (15)$$

Case 1.—Entire population derived from cross of $AABB \times aabb$. First generation 100% $\frac{AB}{ab}$.

$s_0 = 1 \quad s_1 = 0 \quad s_2 = {}^1/_2(1-2c) \quad s_3 = {}^1/_2(1-2c)(1-c)$

$t_0 = -1 \quad t_1 = 0 \quad t_2 = 0 \quad t_3 = \frac{\sigma}{2}(1-2c)$

$r_0 = 1 \quad r_1 = 1-2c \quad r_2 = (1-2c)(1-c) \quad r_3 = (1-2c)(1-c)^2$

$k_1 = 1-2c \quad k_2 = {}^1/_2(1-2c)^2 \quad k_3 = {}^1/_2(1-2c)^2(1-c)$

$$r_\infty = \frac{\sigma(1-2c)}{2c + \sigma + 2c\sigma}, \quad 2x_\infty = \frac{c(1+2\sigma)}{2c + \sigma + 2c\sigma} \quad \text{where } \sigma = \frac{N_m + N_f}{4N_mN_f} \qquad (16)$$

Case 2.—Mixture of $\frac{N_m}{2}$ males, $\frac{N_f}{2}$ females of type $AABB$ with $\frac{N_m}{2}$ males and $\frac{N_f}{2}$ females of type $aabb$, followed by random mating (in each group).

$s_0 = 1 \quad s_1 = {}^1/_2 \quad s_2 = {}^1/_2(1-c)$

$t_{(mm)_0} = \frac{-1}{N_m - 1} \quad t_1 = 0 \quad t_2 = {}^1/_2\sigma$

$t_{(ff)_0} = \frac{-1}{N_f - 1} \quad r_1 = 1-c \quad r_2 = (1-c)^2$

$t_{(mf)_0} = 0 \quad k_1 = {}^1/_2(1-2c) \quad k_2 = {}^1/_2(1-2c)(1-c)$

$$r_0 = 1 \qquad k_3 = \frac{1-2c}{2}\left[(1-c)^2 - \frac{\sigma}{2}\right]$$

$$r_\infty = \frac{\sigma}{2c+\sigma+2c\sigma}, \quad 2x_\infty = \frac{c(1+\sigma)}{2c+\sigma+2c\sigma} \quad \text{where } \sigma = \left(\frac{N_m+N_f}{4N_mN_f}\right) \tag{17}$$

There are several special cases of interest. (*A*) Population equally divided between males and females $N_m = N_f = \frac{N}{2}, \sigma = \frac{1}{N}$

Case 1 $\left(\text{from } \frac{AB}{ab}\right)$,

$$r_\infty = \frac{1-2c}{2(N+1)c+1} \qquad 2x_\infty = \frac{(N+2)c}{2(N+1)c+1} \tag{18}$$

Case 2 $\left(\text{from } \frac{AB}{A\bar{B}} + \frac{ab}{a\bar{b}}\right)$,

$$r_\infty = \frac{1}{2(N+1)c+1} \qquad 2x_\infty = \frac{(N+1)c}{2(N+1)c+1} \tag{19}$$

In brother-sister mating $\left(\text{from } \frac{AB}{ab} \times \frac{AB}{ab}\right)$, $N = 2$ in Case 1 above. The results check with those of Haldane and Waddington

$$r_\infty = \frac{1-2c}{1+6c}, \qquad 2x_\infty = \frac{4c}{1+6c} \tag{20}$$

(*B*) One male, indefinitely many females, ($N_m = 1$, $N_f = \infty$, $\sigma = {}^1/_4$)

Starting from $\frac{AB}{ab}$:

$$r_\infty = \frac{1-2c}{1+10c} \qquad 2x_\infty = \frac{6c}{1+10c} \tag{21}$$

The results are the same as with lines of 2 males, 2 females starting from $\frac{AB}{ab}$. Decrease in heterozygosis also takes place at the same rate.

N Hermaphrodites, no self-fertilization

$$s = {}^1/_2(r' + t')$$

$$t = \frac{1}{N}s' + \frac{N-1}{N}t' = t' + \frac{1}{N}(s' - t')$$

$$r = s + k$$

$$r = r' - c(r' - t').$$

These are identical with the results from the preceding case for populations of N equally divided between males and females $\left(\sigma = \frac{1}{N}\right)$. The general formula for r is consequently the same and the formulae for r_∞ and $2x_\infty$ also agree both in the case of origin from heterozygotes and in origin from population mixture.

N Hermaphrodites, random union of gametes including self-fertilization

In this somewhat arbitrary case the formula for t differs slightly from that above.

$$t = \frac{1}{N}s + \frac{N-1}{N}t' = t' + \frac{1}{2N}(r' - t')$$

$$k = k'\left(1 - c - \frac{1}{2N}\right)$$

$$r_\infty = 1 - \left(\frac{4Nc}{(1-2c)(2Nc-1)}k_2\right) \qquad (22)$$

Case (1) from 100 % $\frac{AB}{ab}$, $r_\infty = \frac{1-2c}{2Nc+1}$ $2x_\infty = \frac{(N+1)c}{2Nc+1}$ (23)

Simplest case $N = 1$ (self fertilization) $r_\infty = \frac{1-2c}{1+2c}$ $2x_\infty = \frac{2c}{1+2c}$ as given before (24)

Case (2) from 50% $\frac{AB}{ab}$ + 50% $\frac{ab}{ab}$

$$r_\infty = \frac{1}{2Nc+1} \quad 2x_\infty = \frac{Nc}{2Nc+1} \qquad (25)$$

Limited Populations with Maximum Avoidance of Consanguineous Mating.—If mating of close relatives is avoided as far as possible in groups of limited size, heterozygosis decreases only about half as rapidly as if mating is at random. There is naturally more chance for random recombination under the former system. To measure this effect, consider first the case of population of four in which all matings are between double first cousins instead of at random.

Let $r = r_{AB}$ for the same gamete
$s = r_{\dot{A}B}$ for two gametes of the same individual
$t = r_{AB}$ for gametes of brothers and sisters
$u = r_{AB}$ for gametes of double first cousins

$s = \frac{1}{2}(r' + u')$ $r = k + s$
$t = \frac{1}{2}(s' + u')$ $r = r' - c(r' - u')$(interpolation)
$u^1 = \frac{1}{2}(t' + u')$ $k = \frac{1-2c}{2}(r' - u')$

$k = \frac{(1-2c)}{2}(k' + {}^1/_2k'' + {}^1/_4k''')$ iteration formula for k

$$\sum_1^n r - \sum_0^{n-1} r = -\frac{2c\sum_1^n k}{1-2c} = r_n - r_0$$
$$\sum_1^\infty k = \left(\frac{8}{1+14c}\right)\left[k_1\left(\frac{1+6c}{4}\right) + k_2\left(\frac{1+2c}{2}\right) + k_3\right] \tag{26}$$

In the case of a population derived wholly from a cross $AABB \times aabb$

$$r_\infty = \frac{1-2c}{1+14c}$$
$$2x_\infty = \frac{16c}{1+14c} \tag{27}$$

In the case of octuple second cousins the iteration formula for k is $\left(\frac{1-2c}{2}\right)(k' + {}^1/_2k'' + {}^1/_4k''' + {}^1/_8k'''')$ and it is easy to show that analogous formulae apply in larger populations in which consanguineous mating is avoided as much as possible. It appears that the general formulae for such a population of size N is as follows

$$r_\infty = \frac{1-2c}{1+2(2N-1)c} \qquad 2x_\infty = \frac{2Nc}{1+2(2N-1)c} \tag{28}$$

Thus with large subgroups, $r_\infty = \frac{1-2c}{4Nc}$ if consanguinity is avoided while it is twice as great, $\frac{(1-2c)}{2Nc}$, if mating is at random after initial cross breeding, and still greater $\left(\frac{1}{2Nc}\right)$ if there is random mating in populations formed by mixture of the two types $AABB$ and $aabb$, in equal numbers.

In any case the limiting correlation is negligible unless the product of size of the breeding populations by the cross-over percentage is low.

Two Sex-Linked Genes in Populations of N_m Males N_f Females.—Formulae can be written by which the correlation between A and B in gametes can be calculated generation after generation.

$s = {}^1/_4(r' + r'' + 2t')$ ($s = r_{AB}$ for gametes of same female)

$t = {}^1/_2(t' + u') + \frac{1}{2N_f}(s' - u')$ ($t = r_{AB}$ for egg and sperm)

$$u = {}^1/_2 t + {}^3/_4 t' - {}^1/_4 t'' + 1/(4N_m)(r'' - 2r' + t'') \quad (u = r_{AB} \text{ for eggs of different females}) \tag{29}$$

$$r = k + s = {}^1/_2(r' + r'') - \frac{c}{2}(r' + t'' - 2t')$$

$$k = \frac{1 - 2c}{4}(r' + r'' - 2t') = (1 - 2c)(s - t')$$

In the special case in which $N_f = 1$ (brother-sister, one male many half-sisters in flocks of birds), $t = {}^1/_2(t' + s')$

$k = k' \frac{(1 - c)}{2} + k'' \frac{(1 - 2c)}{4}$ as in the case of autosomal genes, parent by offspring.

$$r_\infty = \frac{3 - 4c}{3(1 + 4c)} \qquad 2x_\infty = \frac{8c}{3(1 + 4c)} \tag{30}$$

These agree with the results obtained by Haldane and Waddington for 2 sex-linked genes under brother-sister mating.

The general conclusions from this analysis are that different pairs of allelomorphs, even in the same chromosome, come to be combined practically at random in any freely interbreeding population of long standing and that there is practically random combination among the genes fixed in different inbreeding subgroups of a population, unless these subgroups are very small or linkage is extremely close. It is only in exceptional cases that a correlation between characters in a natural population should be attributed to linkage.

[1] Robbins, R. B., *Genetics*, **3**, 375–389 (1918).
[2] Weinberg, W., *Z. ind. Abst. u. Vererb.*, **1**, 277–330 (1909).
[3] Haldane, J. B. S., and Waddington, C. H., *Genetics*, **16**, 357–374 (1931).
[4] Wright, S., *Ibid.*, **6**, 111–178 (1921).
[5] Wright, S., these PROCEEDINGS, **19**, 411–420 (1933).

HYÆNODONTIDÆ OF THE UPPER EOCENE OF CALIFORNIA

BY CHESTER STOCK

BALCH GRADUATE SCHOOL OF THE GEOLOGICAL SCIENCES, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated February 17, 1933

Since the announcement[1] of the discovery of upper Eocene mammals in the Sespe deposits north of the Simi Valley, California, excavations have been continued in these beds, particularly at Locality 150 C.I.T. Vert. Pale., by field parties of the California Institute. Recently an opportunity has been afforded the writer to compare a number of the mammalian types found here with related forms known from the late Eocene and Oligocene. This has permitted a more precise determination of the genera and species than was attempted in the original statement. It is planned therefore to present in this and in subsequent papers brief descriptions of individual types, reserving a fuller discussion for a monographic treatment of the Sespe faunas.

CREODONTA

Family Hyænodontidæ

Hyænodon vetus, n. sp.

Type Specimen.—Skull with upper dentition, No. 1243 C.I.T. Coll. Vert. Pale., plate 1, figure 1, textfigures 1 and 2.

Locality.—Sespe Upper Eocene, north of Simi Valley, Ventura County, California; Locality 150 C.I.T. Vert. Pale.

Specific Characters.—Larger than *Hyænodon crucians* and approaching *H. montanus* and *H. cruentus* in size. Skull slender with postorbital constriction relatively more pronounced than in Oligocene species. Depression of frontal surface between temporal ridges broad and relatively deep. Union of temporal ridges with sagittal crest slightly in advance of level of deepest portions of postorbital constriction. Posterior position of postglenoid processes like that in *crucians;* processes situated slightly farther back than in *paucidens* and in *mustelinus* but not so far back as in *cruentus* and in *horridus*. Backward extent of tooth-row not so great as in American Oligocene species. Anterior premolars closely spaced but not crowded. $P\underline{1}$ with two well developed roots.

Remarks.—The beautifully preserved skull of *Hyænodon vetus* represents the first specifically determinable material of this genus to be recorded from an American Tertiary horizon older than the White River. No. 1243 exhibits several characters in which it may be regarded as more primitive than representatives of the genus in the American Oligocene. The

position of the posterior choanae varies in skulls of the latter. In *H. vetus* the floor to the postnarial passage apparently does not extend backward quite so far as in the Oligocene hyænodons. In No. 1243 the medial walls of the palatines are virtually in contact, in part due perhaps to lateral crushing which has affected this portion of the postnarial passage, but a narrow cleft appears to prevail between them extending forward farther than in the Oligocene forms. Likewise the pronounced postorbital constriction and possibly the lack of a noticeable backward extension of the tooth-row may be regarded as characters in which *H. vetus* is less advanced.

Hyænodon (Protohyænodon) exiguus, n. subgen. and n. sp.

Type Specimen.—A maxillary fragment with $P\underline{3} - M\underline{2}$, No. 928 C.I.T. Coll. Vert. Pale., plate 1, figures 3, 3*a*.

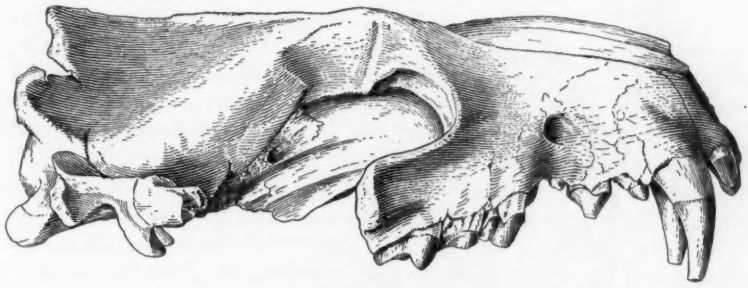


FIGURE 1

Hyænodon vetus, n.sp. Skull, No. 1243, lateral view; × 1/2. C.I.T. Coll. Vert. Pale. Sespe Upper Eocene, California.

Paratypes.—Maxillary fragment with $P\underline{4}$, $M\underline{1}$ and $M\underline{2}$ and a ramus with teeth, No. 1140 C.I.T. Coll. Vert. Pale., plate 1, figures 2, 2*a* and 4, 4*a*.

Locality.—Sespe Upper Eocene, north of Simi Valley, Ventura County, California; Locality 150 C.I.T. Vert. Pale.

Subgeneric Characters.—$P\underline{4}$ with distinct inner cusp. $M\underline{1}$ with tiny protocone. Slight bifurcation of tip of principal antero-external cusp in $M\underline{2}$.

Specific Characters.—Distinctly smaller than *Hyænodon mustelinus* and comparable in size to individuals of *H. vulpinus* from the Phosphorites of Quercy. Approaches closely in size the American species ?*Pseudopterodon minutus*. Width of $P\underline{4}$, relative to length, greater than in latter species.

Remarks.—*Protohyænodon exiguus*, in the several characters stated above, is distinctly more primitive than typical *Hyænodon*. In the former the inner lobe of $P\underline{4}$ projects inwardly farther than in the Oligocene species

and a distinct cusp is present. This region is better developed in the Simi form than in ?*Pseudopterodon minutus*. The inner face of the principal cusp facing the lobe is concave. In the American Oligocene species, where the inner cusp is not so well defined or is absent, the inner surface slopes upward from the tip of the principal cusp to the side of the lobe. In other words, the lobe supporting a cusp of its own serves less as a buttress to the principal cusp in *P. exiguus* than in the Oligocene forms. In $M\underline{1}$ of *P. exiguus* a tiny but distinct protocone is present and is defined from the inner wall of the antero-external cusp by a small furrow. The presence of an antero-internal cusp has been noted by Matthew in $M\underline{1}$ of ?*Pseudopterodon minutus*. In the Sespe species an antero-internal cusp is suggested in $M\underline{2}$, but is not so well marked as in $M\underline{1}$. At the tip of the principal cusp in $M\underline{2}$ of the type specimen of *P. exiguus* is a slight indentation of the enamel surface on both the external and internal sides, separating it into two minute parts. The separation of the tip of this cusp into two parts is to be noted in $M\underline{1}$ of the American Oligocene species but I have not observed it in $M\underline{2}$.

FIGURE 2

Hyænodon vetus, n.sp. Skull, No. 1243, dorsal view; × 1/2. C.I.T. Coll. Vert. Pale. Sespe Upper Eocene, California.

$M\underline{3}$ is absent, in which respect *P. exiguus* is like *Hyænodon*. In the paratype (Fig. 4*a*) minute depressions or perforations have been noted to the inner side of the posterior part of $M\underline{2}$, but it appears unlikely that these represent the remnants of the alveolus for a third molar. Moreover, the width of the maxillary wall to the inner side of $M\underline{2}$ in the type specimen is certainly not great enough to accommodate a third molar in that specimen.

Among the American species, *H. minutus*, described by Douglass[2] from the Pipestone beds of Montana, appears to show closest relationship to *Protohyænodon exiguus*. The maxillary fragment with a portion of the superior dentition later recorded by Matthew[3] from these deposits furnished

the basis for assigning Douglass' species tentatively to *Pseudopterodon*. Some doubt prevails regarding the validity of Schlosser's type for, according to Matthew, Scott has expressed belief that *Pseudopterodon* is based on deciduous teeth of *Hyænodon*. Martin[4] likewise regards this genus as having been based on the milk dentition of *Hyænodon*. In each of the upper carnassial teeth of *Pseudopterodon* the two cusps forming the antero-external region of the crown are widely spaced and their separation is much better defined than in our form. *Protohyænodon exiguus* differs from Matthew's specimen in larger size, in the proportions of $P\underline{4}$ and in the presence of a distinct inner cusp in this tooth.

Among European species, *Hyænodon vulpinus* from the Phosphorites (Quercy) resembles our type in size. This species evidently displays noteworthy differences in size. In the upper dentition of *vulpinus* $P\underline{3}$ is elongate anteroposteriorly, for the region of the crown in front of the principal cusp is extended farther forward than in American species. $P\underline{4}$ in *vulpinus* (No. 11462 Prin. Univ. Coll.) has a distinct inner cuspule. *P. exiguus* is more like *H. vulpinus* in this character than are the American Oligocene species. The protocone in $M\underline{1}$ of *vulpinus* tends to be more prominent than in the latter, in which respect our form is again more like the former.

In at least some specimens of *vulpinus* the anterior basal tubercle of $P\bar{4}$ is more prominently developed than in No. 1140 from the Sespe. In *P. exiguus* a tubercle is absent but a small cingulum is present at the antero-internal base of the principal cusp. The most noteworthy difference in the lower dentition occurs, however, in $M\bar{3}$. Relative to the length of either $M1$ or $M2$, this tooth has a greater length in *H. vulpinus* than in *P. exiguus*.

Pterodon californicus, n .sp.

The genus *Pterodon*, now recorded for the first time from an early Tertiary horizon in North America, is represented by incomplete rami of lower jaws of three individuals. In two specimens, Nos. 933 and 1213 C.I.T. Coll. Vert. Pale., the deciduous dentition is present with $M\bar{1}$ and $M\bar{2}$. No. 941, however, represents an older individual in which the permanent dentition includes not only the first and second molars, but also the third and fourth premolars.

All the specimens from the Simi Sespe show the characteristic structure of the molars in which *Pterodon* differs from *Hyænodon*. In size of talonid and size of paraconid blade in the molars these specimens are like representatives of the *Pterodon* group and differ from *Hyænodon*.

Type Specimen.—Fragment of left ramus with $P\bar{3} - M\bar{2}$, No. 941 C.I.T. Coll. Vert. Pale., plate 1, figure 6.

Paratypes.—Two left rami with deciduous teeth and first and second molars, No. 1213 (plate 1, Fig. 5) and No. 933.

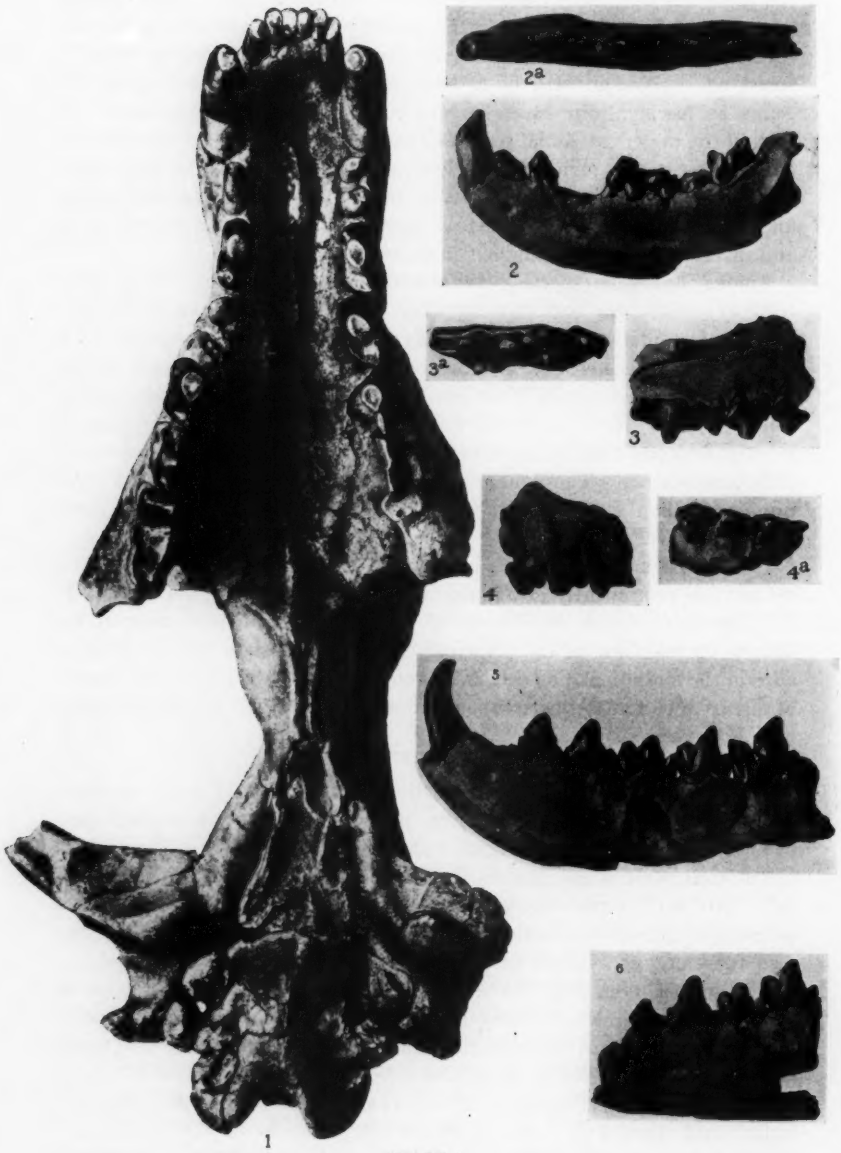


PLATE 1

(Description on opposite page.)

Locality.—Sespe Upper Eocene, north of Simi Valley, Ventura County, California; Locality 150 C.I.T. Vert. Pale.

Specific Characters.—*Pterodon californicus* approaches most closely in size the species *P. leptognathus* from the Fayûm, but is distinctly smaller. $P\bar{3}$ and $P\bar{4}$ distinctly more robust and anterior molars relatively smaller than in *P. leptognathus*. Decidedly smaller than *Hemipsalodon grandis*.

Remarks.—The three specimens available show the characters of the lower deciduous dentition and the acquisition of the permanent teeth. The youngest specimen, in terms of development of the dentition, is No. 933. In this jaw the milk incisors, canine and milk premolars are present. The milk incisors are crowded in transverse row with $DI\ \bar{2}$ displaced slightly backward. $DI\ \bar{2}$ and $DI\ \bar{3}$ are of subequal size. The crown of a permanent incisor is appearing behind the deciduous teeth and immediately adjacent to the symphyseal surface. $Dp\ \bar{1}$ and $Dp\ \bar{2}$ are two-rooted, with the anterior root considerably more reduced than the posterior. With reduction in size of the anterior root and procumbent position of crown, the posterior root in $Dp\ \bar{1}$ extends obliquely backward. The single alveolus for $Dp\ \bar{1}$ in No. 1213 may have accommodated not only the principal root but also the reduced forward fang. The crown of $Dp\ \bar{1}$ extends very close to the canine, flanking the lower posterior border of this tooth. The posterior root of $Dp\ \bar{2}$ likewise projects backward in the jaw. In this tooth the highest point reached by the principal cusp is situated farther posteriorly than that of $Dp\ \bar{1}$, and the crown resembles more that of $Dp\ \bar{3}$.

The structure of $Dp\ \bar{3}$ is like that of $P\bar{4}$, although the crown is more compressed than in the latter tooth. The principal cusp, which does not stand so high as in $P\bar{4}$, is slightly recurved. A tubercle of minute size is present at the antero-internal side of the base of the principal cusp, while a much larger posterior basal cusp is also present. $Dp\ \bar{4}$ resembles somewhat the first molar. It differs, however, from the latter tooth in a number of characters. $Dp\ \bar{4}$ is longer and more compressed, the paraconid and protoconid are not so blade-like and the paraconid shear shows less obliquity with reference to the anteroposterior axis of the tooth than in the first molar. A distinct talonid is present. In specimen No. 1213 the partly developed permanent premolar is seen in the jaw beneath $Dp\ \bar{4}$.

In the type specimen, No. 941, considerable disparity in size exists

---

DESCRIPTION OF PLATE 1

Figure 1, *Hyænodon vetus*, n .sp. Type specimen, No. 1243. Figures 2, 2a, 3, 3a, 4, 4a, *Hyænodon* (*Protohyænodon*) *exiguus*, n. subgen. and n. sp.; figures 3, 3a, type specimen, No. 928; figures 2, 2a and 4, 4a, paratypes, No. 1140. Figures 5 and 6, *Pterodon californicus*, n. sp.; figure 6, type specimen, No. 941; figure 5, paratype, No. 1213. All figures $\times\ ^3/_4$.

California Institute of Technology Collections. Sespe Upper Eocene, California.

between $P\bar{3}$ and $P\bar{4}$. $P\bar{3}$ possesses a broad base with principal cusp anterior in position. No well developed posterior basal tubercle is present. In $P\bar{4}$ the principal cusp stands high and a posterior basal cusp was presumably present although the enamel surface is marred in this region. The crown of this tooth does not tilt backward. The obliquity of the paraconid shear in the molars, relative to the fore and aft axis of the individual tooth, does not appear to be as great as that in *P. leptognathus* and approaches more that seen in *P. dasyuroides*. The posterior edge of the protoconid reaching upward for nearly half the distance between the level of the talonid and the tip of the cusp projects noticeably backward and can be readily noted in the illustration (plate 1, Fig. 5). In the more adult specimen, No. 941, this portion of the edge is worn. A characteristic enamel ridge extending for a short distance longitudinally on the antero-external surface of the base of the paraconid marks laterally the border of the area against which the tooth in front is pressed.

The posterior end of the mandibular symphysis in No. 941 reaches a point below the anterior end of $P\bar{4}$. In this specimen, also, the large posterior mental foramen lies below the anterior root of $P\bar{4}$. A distinct groove extends from this opening anteriorly, reaching a point below the anterior end of $P\bar{3}$. The very large ramus of the mandible of *Hemipsalodon grandis*, described by Cope from the White River Oligocene of Swift Current Creek, Canada, is widely removed specifically from the Californian form and is a later type.

[1] Stock, C., *Proc. Nat. Acad. Sci.*, **18**, 518–523 (1932).
[2] Douglass, E., *Trans. Amer. Philos. Soc.*, **20**, 255 (1901).
[3] Matthew, W. D., *Bull. Amer. Mus. Nat. Hist.*, **19**, 208–209 (1903).
[4] Martin, R., *Rev. Suisse de Zoöl.*, **14**, 575 (1906).

THE EFFECT OF HELIUM ON THE CONTINUOUS AND SECONDARY SPECTRA OF HYDROGEN

BY A. S. ROY*

UNIVERSITY OF ST. ANDREWS

Communicated March 6, 1933

The main factors in the distribution of intensity among the lines of a spectrum are: (1) electrical conditions of excitation, (2) pressure of the gas in the discharge tube and (3) presence of impurities.

It has long been known that the effect of the addition of the inactive gases to substances emitting spectra in certain cases modifies to a great extent their normal spectra. In some cases entirely new spectra are excited. Many remarkable results have been found by previous investigators using hydrogen. Liveing and Dewar[1] in 1900 observed that the higher members of the Balmer series were excited in mixtures of hydrogen and the rare gases, and more recently Merton and Nicolson[2] have examined the same mixtures. Later Merton[3] examined the secondary spectrum when excited in mixtures of helium and a trace of hydrogen at a total pressure of 40 mm. of Hg. The author in a paper in publication (*Philosophical Magazine*) has found that the Balmer lines are also excited more strongly in mixtures containing an excess of oxygen. The present paper is an account of part of the investigation which has been made in the Physical Laboratory of The University of St. Andrews on the effect of foreign gases on gaseous spectra; the effect of the addition of helium to hydrogen was investigated by the writer in an attempt to find a possible explanation of the observed changes in the relative intensity of the lines of the secondary spectrum of hydrogen.

Apparatus.—Two Hilger discharge tubes were mounted so that either could be brought into line with the spectrograph and so the spectra from each tube could be photographed in juxtaposition. The hydrogen was generated by the electrolysis of a dilute solution of sodium hydroxide and dried in the usual manner by means of a system of drying agents. The helium was supplied by the British Oxygen Company and contained a small trace of neon. The spectrograph used was the Hilger E56 having a glass optical system.

The pressure of hydrogen in the discharge was always approximately 0.1 mm. of Hg; pressures of helium up to 20 mm. of Hg were used but the best results were obtained at pressures of 10 mm. of Hg. At higher pressures than this no further change in the relative intensity of the lines took place and at these pressures very long exposures were necessary to bring out the lines of the secondary spectrum of hydrogen.

The secondary spectrum of hydrogen was examined from H_α to H_γ and

the intensities compared with those given in the tables of Gale, Monk and Lee,[4] and also with the intensities given by Kapuscinski and Eymer.[5]

Results and Discussion.—The lines of the secondary spectrum of hydrogen were effected by the addition of helium in different ways, as was first shown by Merton. The lines fall into three groups: (1) lines which are enhanced, (2) lines which show no appreciable alteration and (3) lines which show an increase in intensity after the addition of helium.

It was found from the examination of the secondary spectrum in the region photographed the lines of the triplet system are, in general, unaffected or increased in intensity, while the lines of the singlet system are generally diminished in intensity. The triplet lines which showed the greater enhancement belonged to the systems recently arranged by Richardson and Davidson[6] and also by Sandeman.[7] The continuous spectrum of hydrogen was markedly increased in intensity in the spectrum of the helium and hydrogen mixture.

An explanation of this enhancement of the lines of the triplet system and decrease in the intensity of the lines of the singlet system relative to their intensities in a pure hydrogen discharge may be given assuming that the helium brings about a limitation of electron velocities.

Brasefield[8] has studied the variation of the lines of the secondary spectrum of hydrogen and he found that the most important factor which influenced changes in intensity was the velocity of the exciting electrons. He observed that there was always a decided difference in the behavior of the singlet and triplet lines. The curves showing the variation of intensity of the lines of the triplet system with the energies of the exciting electrons had no maxima in the range of velocities used but seemed to be approaching one slightly above 19 volts. On the other hand the singlet lines showed a decided maximum of intensity, between 30 and 40 volts. Massey and Mohr[9] have now verified Brasefield's experimental results; from theoretical considerations they obtained curves showing the probability of excitation with the velocity of the exciting electrons. They found that in a two electron system the triplet states differ from the singlet states in their behavior toward electron impact. The probability of exciting a triplet state had a sharp maximum for electron energies just above the excitation potential, whereas in the case of singlet states the maximum probability is not sharply defined and occurs at energies greater than the excitation potential of the lines.

The above results may now be applied in the case of the helium and hydrogen mixtures. Helium has metastable states at 19.73 and 20.51 volts and an ionization potential at 24.47 volts. Since the pressures of helium were high it is extremely probable that the helium brought about a limitation of the electron velocities. Therefore, since the excitation functions of the lines of the two systems of hydrogen are different at lower

electron velocities compared with that at higher velocities, a redistribution of the intensity of the lines of the secondary spectrum would take place. The lines of the triplet system would therefore be enhanced and the lines of the singlet system diminished in intensity.

As has been pointed out above, the continuous spectrum of hydrogen was very much stronger in the mixtures of helium and hydrogen than in pure hydrogen. This observation strengthens the fact that the probability of exciting the triplet states is greater than the singlet states when helium is present. There have been many theories to explain the continuous spectrum of hydrogen; the most recent theory put forward to explain its excitation and one which is now generally accepted is that due to Winans and Stueckelberg.[10] They suggested that the continuous spectrum was due to transitions from the higher excited triplet states to the normal unquantized triplet level—$1^3\Sigma$. Now since we have shown by the method of limitation of velocities that the triplet states would be more probably excited than the singlet states in the mixtures of helium and hydrogen it follows that the continuous spectrum would necessarily be enhanced.

A number of the bands of the secondary spectrum of hydrogen, as arranged by Richardson, were particularly examined in order to study the relation between the change in intensity with change in rotation quantum number. Although Richardson's[11] previous results were verified and somewhat extended, no theory can at present be put forward to explain the apparent regular variation in the intensity with increase in rotation quantum number. At present the writer is continuing the investigation of the excitation of rotation states of electronic bands, with special attention to their excitation by impacts of the second kind, at the University of Michigan.

Conclusion.—An examination of the secondary spectrum and continuous spectrum of hydrogen has been made in mixtures of helium and hydrogen. The lines of the triplet system were found to be, in general, increased in intensity relative to the lines of the singlet system; at the same time the continuous spectrum was very much stronger. This has been explained by the limitation of electron velocities by the helium, which would increase the probability of exciting the triplet states of hydrogen.

No theory can as yet be put forward to explain the change in intensity of the lines in the bands of the secondary spectrum with change in rotation quantum number.

* COMMONWEALTH FELLOW at the University of Michigan.

[1] Liveing and Dewar, *Proc. Roy. Soc.*, **67**, 467 (1900).

[2] Merton and Nicolson, *Phil. Trans.*, **217**, 237 (1918).

[3] Merton, *Ibid.*, **222**, 369 (1922).

[4] Gale, Monk and Lee, *Astrophys. J.*, **67**, 89 (1928).
[5] Kapuscinski and Eymer, *Proc. Roy. Soc.*, **122**, 58 (1929).
[6] Richardson and Davidson, *Ibid.*, **131**, 658 (1931).
[7] Sandeman, *Ibid.*, **138**, 395 (1932).
[8] Brasefield, *Phys. Rev.*, **32**, 361 (1928).
[9] Massey and Mohr, *Proc. Roy. Soc.*, **132**, 605 (1931).
[10] Winans and Stueckelberg, *Proc. Nat. Acad. Sci.*, **14**, 867 (1928).
[11] Richardson, *Proc. Roy. Soc.*, **111**, 721 (1926).

K SERIES SPECTRUM OF TUNGSTEN

BY J. C. HUDSON AND H. G. VOGT

RESEARCH LABORATORY OF PHYSICS, HARVARD UNIVERSITY

Communicated March 11, 1933

The K series x-ray spectrum of tungsten has been investigated by a large number of previous workers, among whom may be mentioned Duane and Shimizu,[1] Siegbahn,[2] Duane and Stenström,[3] Rechou,[4] Cork,[5] Stephenson and Cork,[6] Seemann[7] and Williams.[8] M. de Broglie[9] in 1920 separated the β doublet into two components, and in 1930 Armstrong[10] secured complete separation of this doublet and also secured a δ line very near the absorption limit. Seemann and Williams each succeeded in separating the β doublet also, though the separation is not complete in either case.

The present note gives some results of photographic measurements of the tungsten K series lines in which both the β and γ doublets are resolved and the wave length of the δ line determined.

Figure 1 is a photograph of the apparatus. The source of x-rays is a General Electric Pyrex tube operated at a constant potential difference of 165 k.v. and a current of 8 m.a. The power plant and the lead box containing the tube can be seen through the door in the background of the photograph. The room containing the tube and power source is completely outside the main building and the ceiling of it is about six feet below the surface of the ground. Temperature changes in the two rooms shown are very small. A calcite crystal (about the center of Fig. 1) diffracts the x-ray beam and the focusing arrangement proposed by Bragg is used, with the distance from the axis of the spectrometer to the photographic film being about six meters. The slit near the tube is about 0.2 mm. wide and 30 mm. in height. In order to reduce absorption by the long air path an evacuated steel pipe four inches in diameter is placed between the first slit and the crystal. The ends of the pipe are closed with brass plates containing narrow slots covered with cellophane and the x-ray beam passes through the slots. The larger pipe in the fore-

ground of the photograph can also be evacuated though this was not done in the present experiments.

In order to secure as great an intensity as possible at the crystal considerable care was taken in adjusting the tube relative to the first slit and the axis of the spectrometer. The tube is mounted so that it can be shifted parallel to its long axis by means of a fine screw and it can be rotated about a vertical axis passing through the center of the focal spot. A point counter[11] with a very narrow slit was set up over the axis of the

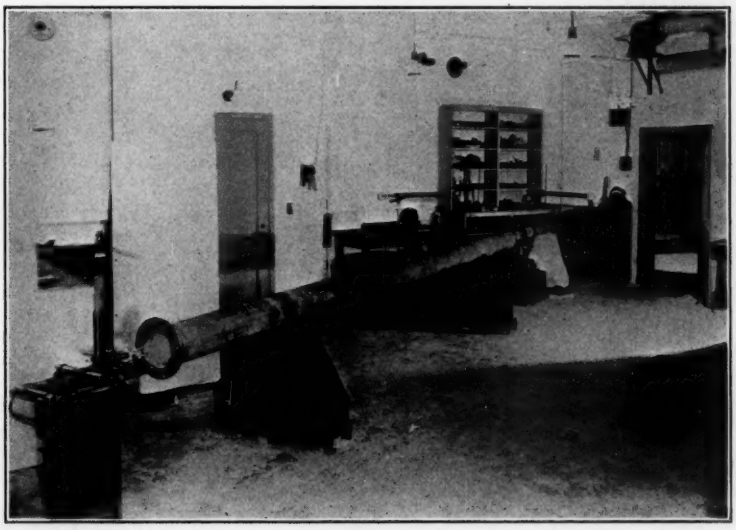


FIGURE 1

spectrometer and the tube adjusted until the counter discharge rate was a maximum.

The crystal used is a specimen of calcite from Montana and the diffracting face is 10 cm. by 7 cm. A rather long exposure of the α lines with the crystal fixed gave fine lines with no indication of multiple lines which are sometimes present due to irregularities on the surface. No steps are visible on the crystal face used in the experiments. All the spectral lines were secured on each photograph used in the final measurements and for each setting of the crystal a lever arm and cam served to oscillate the crystal through an arc of five minutes. Thus both α lines are secured with one setting and the β and γ doublets require one setting each. The δ line is secured at the same setting as the γ doublet.

Figure 2 is a reproduction of one of the films without magnification

and shows the two α lines on the right, then the β doublet, clearly resolved, and the two γ lines. The δ line is rather faint and does not reproduce well.

The wave lengths of the β, γ and δ lines are calculated from the values of α_2 and α_1 given by Duane and Stenström.[3] These values are 0.21341 Å and 0.20860 Å, respectively, assuming the value of d of calcite for first order reflection to be 3.028 Å. Birge[12] gives 3.0279 Å as the value of d for first order reflection and this is so near the value used by Duane and

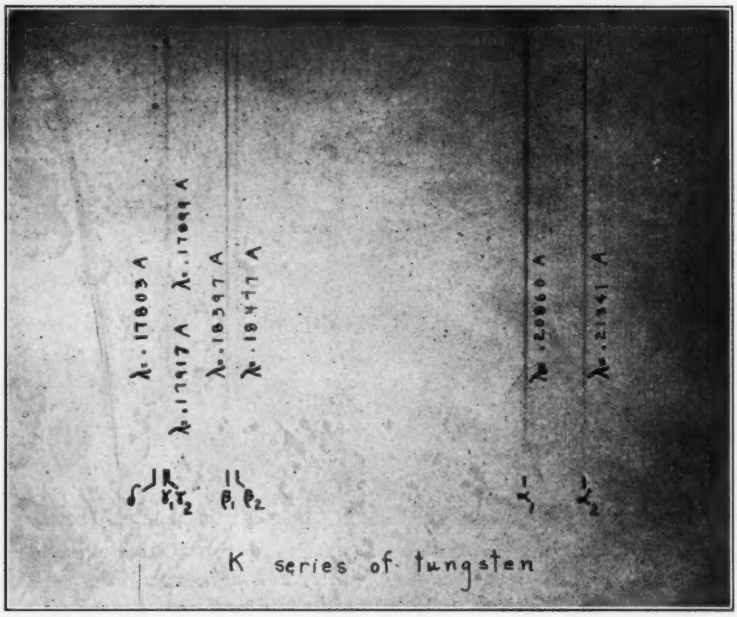


FIGURE 2

Stenström that no correction has been made in their values of wave lengths. If we average the values of the wave lengths of the α lines (changing the wave lengths to the values they have if d for first order reflection is taken as 3.028 Å) as given by Duane and Stenström,[3] Siegbahn,[2] Stephenson and Cork[6] and Rechou[5] we secure values differing from the ones used by not more than 0.00001 Å. The value of α_2 given by Williams[8] is appreciably smaller than that secured by most other experimenters. Using his values of α_2 and α_1 would of course shift our measured lines toward shorter wave lengths.

In making the measurements for determining the wave lengths of the

β and γ doublets and the δ line two different photographs were used, each photograph showing all seven lines. The distances between the various lines were determined by a comparator which read directly to 0.001 mm. All the lines were measured by each of three observers and the results averaged. As an indication of the separation of the lines of the photographs the distance between the α lines was 9.548 mm. The values of the wave lengths are given in the following table, the first line being the values secured in the present experiments. The values for the α lines are taken from the paper by Duane and Stenström.[3] The second line is the present averaged values for the β and γ doublets on the basis that the shorter wave length line has twice the intensity of the longer component. The notation used is an extension of that adopted by the Braggs, the subscript 1 denoting the shorter wave length component of the doublet in each case.

The remaining values in the table are those of other observers for comparison. Only an average value of the γ doublet is available for comparison, as it has not been previously resolved so far as we know. All the values of wave lengths are given on the basis of d for calcite first order reflection being 3.028 Å.

	α_2	α_1	β_2	β_1	γ_2	γ_1	δ
Present	0.21341	0.20860	0.18477	0.18397	0.17917	0.17899	0.17803
				0.18423		0.17905	
Duane and Stenström	0.21341	0.20860		0.18420		0.17901	
Siegbahn	0.21345	0.20878		0.18431		0.17934	
Stephenson and Cork	0.21338	0.20855		0.18416		0.17892	
Seemann			0.18480	0.18399			
Rechou	0.21341	0.20843		0.18410		0.17894	
Williams	0.21330	0.20849	0.18469	0.18391		0.17900	

The δ line is rather faint on our photographs and lies very near the absorption limit. Such a line has been reported for Mo by Leide,[13] and by Duane[14] and for various elements by Ross.[15] Some photographs taken recently in Siegbahn's laboratory also show this line for Mo. Copies of these photographs were very kindly sent by Professor Siegbahn.

This work is being continued in the x-ray rooms of the Research Laboratory of Physics.

[1] Duane and Shimizu, *P. R.*, **14,** 67 (1919).
[2] Siegbahn, *P. M.*, **38,** 639 (1919).
[3] Duane and Stenström, *Proc. Nat. Acad. Sci.*, **6,** 477 (1920).
[4] Rechou, *C. R.*, **180,** 1107 (1925).
[5] Cork, *P. R.*, **25,** 197 (1925).
[6] Stephenson and Cork, *Ibid.*, **27,** 138 (1926).
[7] Seemann, *Zeits. Phys.*, **73,** 87 (1931).
[8] Williams, *P. R.*, **40,** 791 (1932).
[9] de Broglie, *C. R.*, **170,** 1053 (1920).
[10] Armstrong, unpublished photographs secured in this laboratory.

[11] Duane, *C. R.*, **150**, 228 (1910).
[12] Birge, *Phys. Rev. Supplement*, July (1929).
[13] Leide, *C. R.*, **180**, 1203 (1925).
[14] Duane, *Proc. Nat. Acad. Sci.*, **18**, 63 (1932).
[15] Ross, *P. R.*, **39**, 536 (1932).

AN ANALYTIC CLOSED SPACE CURVE WHICH BOUNDS NO ORIENTABLE SURFACE OF FINITE AREA

BY JESSE DOUGLAS

DEPARTMENT OF MATHEMATICS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Communicated March 1, 1933

The author has recently given several different examples of contours in space which bound no surface of finite area. One of these (found in collaboration with Ph. Franklin) was a skew polygon of a denumerable infinity of sides.[1] Another was constructed by winding a thread more and more times about the successive tori of a certain infinite sequence, shrinking to a point as limit.[2]

When one inquires what it is in the nature of this second example that produces the non-finite-area-bounding property, it appears very soon that it is because of the presence in the contour of a spiral whose coils contract with a sufficient order of slowness—the spiral winds 2^n-times about the nth torus before passing to the $(n+1)$th, of half the meridian area.

The present note applies this observation to the construction of a contour defined by simple analytic formulae, such that every orientable surface—of whatever topological structure otherwise—bounded by this contour has infinite area. Whether there exists a non-orientable (one-sided) surface of finite area bounded by the contour is left an open question for the present. Formulae involving the functional $A(g)$, introduced by the author into the problem of Plateau, indicate that, anyway, there cannot be any such of finite connectivity.

The example to be given is in the form of a spiral, analytic everywhere except at the pole toward which its coils contract, where, however, it still has a definite tangent. This spiral may be considered as simple among space curves as, for instance, the reciprocal spiral, $r = \frac{1}{\theta}$, among plane curves. The simplicity and near to hand character of the example show how futile are the pretensions to completeness of any treatment of the Plateau problem that restricts itself to the case of a contour capable of bounding a surface of finite area.

Construction.—Using cylindrical coördinates (r, θ, z), we begin by constructing on the unit cylinder, $r = 1$, the spiral

$$z = \theta^{1/5}, \quad r = 1.$$

The desired example is obtained immediately by taking the inverse of this spiral with respect to the unit sphere.

The resulting curve Γ lies on the torus (with zero opening) generated by revolving about the z-axis the circle of unit diameter in the xz-plane tangent to this axis at the origin. It has the form of a spiral whose coils contract to the origin, which is approached from both sides, above and below—so that the curve is closed.

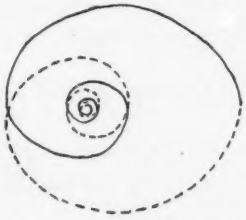
The equations of Γ in spherical polar coördinates (ρ, θ, φ) are easily found to be

$$\theta = \tan^5 \varphi, \quad \rho = \cos \varphi, \tag{Γ}$$

the second of these defining the torus.

The first of its equations, with $-\frac{\pi}{2} < \varphi \leqq \frac{\pi}{2}$, defines Γ as a one-one continuous image of the circle represented, in any meridian plane, by its second equation—which proves that Γ is a Jordan space curve.

Proof of the Infinite Area Property.—The projection of Γ on the xy-plane is a spiral Γ' coiling asymptotically around the origin in two ways, clockwise and counterclockwise, as indicated in the figure.



A good idea of the actual form of Γ may be had by imagining this spiral as a spring, of which the full traced arc is pulled upward, and the dotted arc pressed downward, the pole and the other junction point of the full and dotted arcs remaining fixed.

In polar coördinates (r, θ) in its plane, the equation of Γ' is quickly calculated to be

$$r = \frac{1}{1 + \theta^{2/5}}. \tag{Γ'}$$

The essential fact to be remarked is that the area bounded by Γ' is infinite:

$$\frac{1}{2} \int_{\Gamma'} r^2 d\theta = + \infty. \tag{1}$$

Naturally, this area is to be considered as a Riemann surface, generated by a rotating radius vector whose initial point remains fixed at the origin, while its terminal point describes Γ' from $\theta = -\infty$ to $\theta = +\infty$. We have asymptotically for $\theta \longrightarrow \pm \infty$:

$$r \sim \theta^{-5/6}, \quad r^2 \sim \theta^{-4/6},$$

and therefore

$$\frac{1}{2}\int_{\Gamma'} r^2 d\theta = \frac{1}{2}\left[5\theta^{1/6}\right]_{-\infty}^{+\infty} = +\infty,$$

as stated in (1).

This remark made, the proof that every orientable surface S bounded by Γ has infinite area can be quickly given. We have—where α, β, γ denote the direction angles of the normal to S—

$$\text{Area of } S = \int\int_S d\sigma \geqq \int\int_S d\sigma \cos\gamma = \int\int_S dx\, dy. \quad (2)$$

Now we can put this surface integral in a form appropriate for the application of Stokes' theorem, that is:

$$\int\int_S dx\, dy = \int\int_S \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\, dy + \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) dy\, dz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) dz\, dx,$$

by writing

$$P = -\frac{1}{2}y, \quad Q = \frac{1}{2}x, \quad R = 0;$$

so that, by Stokes' theorem,

$$\int\int_S dx\, dy = \int_\Gamma \frac{1}{2}(x\, dy - y\, dx).$$

Since only the x and y coördinates figure, the last integration may evidently as well be over the xy-projection Γ' of Γ:

$$\int\int_S dx\, dy = \int_{\Gamma'} \frac{1}{2}(x\, dy - y\, dx) = \frac{1}{2}\int_{\Gamma'} r^2 d\theta = +\infty.$$

Therefore, by (2),

$$\text{Area of } S = +\infty.$$

It is evident that this property of bounding no orientable surface of finite area depends solely on the condition (1) for the orthogonal projection of the contour on some plane, a remark which enables us to multiply indefinitely examples of contours with the stated property.

[1] J. Douglas and Ph. Franklin, "A Step-Polygon of a Denumerable Infinity of Sides Which Bounds No Finite Area," these PROCEEDINGS, **19**, 189–191 (1933).

I take this occasion to correct the following misprint in that paper: p. 189, line 16, read "$1 - \frac{1}{2^{n-1}}$" instead of "$\frac{1}{2^{n-1}}$."

[1] J. Douglas, "A Jordan Space Curve Which Bounds No Finite Simply Connected Area," these PROCEEDINGS, **19**, 269–271 (1933).

The restriction to simply connected surfaces was only for simplicity; actually, the infinite area property applies to all orientable surfaces bounded by the contour.

REGULAR DIFFERENTIAL SYSTEMS OF THE FIRST ORDER[1]

BY JOSEPH MILLER THOMAS

DEPARTMENT OF MATHEMATICS, DUKE UNIVERSITY

Communicated March 13, 1933

1. Consider a system of partial differential equations of the first order solved for certain of the derivatives of the unknowns u:

$$R_{i\alpha} \equiv \frac{\partial u_\alpha}{\partial x_i} - F_{i\alpha} = 0.$$

Place the equation $i\alpha$ in the ith row and αth column of an n by r lattice. The result will be called the *array* of the system.

If the independent variables and the unknowns can be and have been numbered so that the empty places form a continuous strip at the right of each row and at the bottom of each column, the system will be called *regular*.

THEOREM 1. *Every regular differential system of the first order is orthonomic.*

This follows at once from the theorem on page 350 of Riquier's treatise,[2] provided we choose the first cotes of all the unknowns equal to zero. A direct proof, simpler than Riquier's, can easily be given also. Assign first cotes 1 and 0 to x and u, respectively. Let the second cote of the independent variable of the lowest row containing a filled place be unity. Proceed upward, increasing the second cote each time a new row is encountered by unity plus the increase in the number of filled places in the row. Give a second cote to the unknowns in like manner, starting at the filled place farthest to the right and proceeding to the left, the increase each time being one greater than the increase in the number of filled places in the column. To the other variables and unknowns give a second cote zero. It then readily follows that the minimum second cote a of the principal derivatives and the maximum second cote b of the parametric derivatives correspond to positions in the array at vertices of the broken line dividing the filled from the empty places, and that $a = b + 2$.

2. When Theorem 1 has once been established, Riquier's existence theorem for orthonomic systems gives immediately

THEOREM 2. *Every passive, regular system of the first order is completely integrable.*

It is not necessary, however, to use the theorem for orthonomic systems in the proof of Theorem 2. In 1884, J. König[3] proved it directly from the Cauchy theorem for the case where the filled portion of the array is m rows. His method of demonstration is, moreover, immediately applicable to the general case. The initial conditions have to be restated and several steps, instead of one, are necessary in eliminating principal derivatives to get the integrability conditions, but these circumstances present no difficulty.

In substance, König's method is as follows. He supposes determined for the unknowns expansions u_α^k which satisfy, for $x_1 = \ldots = x_k = 0$, all the equations below the kth row in the array. He then puts $x_1 = \ldots = x_{k-1} = 0$ in the kth row, and by Cauchy's theorem finds for the resulting equations solutions u_α^{k-1} which reduce to u_α^k for $x_\kappa = 0$. The integrability conditions connecting the kth row with those below it furnish a simple proof that u_α^{k-1} satisfy the equations below the kth row for all values of x_k. Hence the desired result is reached by induction on k.

The principle underlying König's method is of far reaching importance in the subject of existence theorems. It is the vital principle, for example, in the demonstration that the developments set up by Riquier for the unknowns of an orthonomic system actually satisfy the system.[4] Likewise it enables Cartan[5] to base his existence theorem for pfaffian systems directly upon Cauchy's theorem. The latter statement can be seen by an examination of the proof given in Goursat's book. It is brought out even more strikingly in a proof which the writer will give later of a generalized form of Cartan's theorem. At the same time it appears that Cartan's solution of a pfaffian system is effected by means of a passive, regular system of the first order.

3. The following can be proved:

THEOREM 3. *Any system of partial differential equations is equivalent under linear transformation of the independent variables to a regular system of the first order.*

It seems likely that any system can be reduced by a finite number of operations to a passive, regular system of the first order, but the writer has not yet settled the question. Whatever the answer may be, the passive regular form is very useful and important because extensive classes of systems can be reduced to it. One such system is

$$\frac{\partial \psi_\alpha^i}{\partial a_\beta} - \frac{\partial \psi_\beta^i}{\partial a_\alpha} = F_{\alpha\beta}^i \quad (i = 1, 2, \ldots, n;\ \alpha, \beta = 1, 2, \ldots, r),$$

where the F's are functions of the ψ's and a's alone and $F_{\alpha\beta}^i = -F_{\beta\alpha}^i$. It

can be thrown into passive, regular form provided the integrability conditions

$$\frac{\partial F^i_{\alpha\beta}}{\partial x^\gamma} + \frac{\partial F^i_{\beta\gamma}}{\partial x^\alpha} + \frac{\partial F^i_{\gamma\alpha}}{\partial x^\beta} = 0$$

are identically satisfied in ψ, a. This type includes the equations of Maurer as a special case.

[1] Presented to the *American Mathematical Society*, March 25, 1932. The proof mentioned in § 2 was obtained while the writer was NATIONAL RESEARCH FELLOW (1927).

[2] *Les Systèmes d'Equations aux Dérivées Partielles*, Paris (1910).

[3] König, J., *Math. Ann.*, **23**, 520–526 (1884).

[4] Thomas, J. M., "Riquier's Existence Theorems," *Ann. Math.*, **2**, 30, 306 (1929).

[5] Cartan, E., *Ann. Ec. Norm.*, **3**, 18, 260 (1901); Goursat, *Problème de Pfaff*, Paris, 365–373 (1922).

NOTE ON A METRICALLY TRANSITIVE SYSTEM

BY W. SEIDEL[1]

DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY

Communicated March 9, 1933

1. The notion of metrically transitive systems, which has come to play a significant rôle in modern dynamical theory through the work of Koopman, von Neumann, Hopf and Birkhoff was first introduced by G. D. Birkhoff and P. A. Smith,[2] who gave an elementary example.[3] A metrically transitive transformation will be defined as follows.[4] Let Ω be a manifold in n-dimensional space of Lebesgue measure $m(\Omega)$ and let $T(P)$ denote some one-to-one transformation which carries every point P of Ω into some point $T(P)$ of Ω. This transformation will be called *metrically transitive* if every set S of points of Ω which remains invariant under T is either of Lebesgue measure 0 or of Lebesgue measure $m(\Omega)$. In a recent investigation of geodesics on closed surfaces of negative curvature Professor G. D. Birkhoff was led to a certain fundamental transformation whose metrical transitivity seemed likely. I have been able to establish the validity of this conjecture. Since this transformation is of some interest apart from its origin, it will be considered here on its own merits.

2. Let $\ldots a_{-2}\, a_{-1}\, a_0\, a_1\, a_2 \ldots$ denote some sequence of the real numbers 0, 1, 2, ..., 9. The symbol $\ldots a_{-2}\, a_{-1} \cdot a_0\, a_1\, a_2 \ldots$, where a dot has been inserted between two consecutive numbers of the sequence, may be represented by the point (x, y), where $x = .a_0\, a_1\, a_2 \ldots$, $y = .a_{-1}\, a_{-2} \ldots$. Thus our manifold becomes the unit square $0 \leqq x \leqq 1$,

$0 \leqq y \leqq 1$. If the dot in the symbol $\ldots a_{-2}\, a_{-1} \cdot a_0\, a_1\, a_2 \ldots$ is moved one place to the right, the symbol becomes $\ldots a_{-2}\, a_{-1}\, a_0 \cdot a_1\, a_2 \ldots$. Denoting the point which represents the first symbol by (x, y) and the point which represents the second symbol by (x', y'), we thus obtain a transformation $T(P)$ of every point (x, y) into a corresponding point (x', y'). Analytically this transformation is defined by the equations[5]

$$(T) \quad x' = 10x - [10x], \qquad y' = \frac{1}{10}\, y + \frac{1}{10}\, [10x].$$

If we exclude points which have at least one rational coördinate from the square, the transformation (T) becomes one-to-one and continuous.

Let S be any measurable set of points of the unit square $0 \leqq x \leqq 1$, $0 \leqq y \leqq 1$ which remains invariant under the transformation (T). It will be shown that then either $m(S) = 0$ or $m(S) = 1$, $m(S)$ denoting the superficial measure of S.

3. We introduce the function $f(x_0)$ which is equal to the linear measure of that part of S which lies on the segment $x = x_0$, $0 \leqq y \leqq 1$. The function $f(x_0)$ is defined and measurable on the whole interval $0 \leqq x \leqq 1$ except perhaps on a set of measure zero.[6] We have

$$f(x_0) = f_1(x_0) + f_2(x_0) + \ldots + f_{10}(x_0), \tag{1}$$

where $f_n(x_0)$ denotes the linear measure of that part P_n of S which lies on the segment $x = x_0$, $\frac{n-1}{10} \leqq y \leqq \frac{n}{10}$. Now consider the segment $x = \frac{x_0}{10} + \frac{n-1}{10}$, $0 \leqq y \leqq 1$. It is seen that this segment is carried by (T) into the above-mentioned segment, so that we find

$$f_n(x_0) = \frac{1}{10}\, f\left(\frac{x_0}{10} + \frac{n-1}{10}\right).$$

Summing for all n and noting (1), we obtain

$$f(x) = \frac{1}{10} \sum_{n=0}^{9} f\left(\frac{x}{10} + \frac{n}{10}\right). \tag{2}$$

Applying formula (2) to each of the terms $\frac{1}{10}\, f\left(\frac{x}{10} + \frac{n}{10}\right)$ and repeating this process p times yields the functional equation

$$f(x) = \frac{1}{10^p} \sum_{n=0}^{10^p-1} f\left(\frac{x}{10^p} + \frac{n}{10^p}\right). \tag{3}$$

4. We proceed to prove that the only bounded measurable solutions $f(x)$ of the system of equations (3), when p is allowed to assume all posi-

tive integral values, are functions which are constant in the interval $0 \leqq x \leqq 1$ save perhaps for a set of measure zero.[7] We observe that if $f(x)$ is a Riemann integrable function, on letting p become infinite, it follows at once from equation (3) that $f(x) = \int_0^1 f(t)dt$. In order to prove this for the case that $f(x)$ is bounded and measurable, let us define $f(x)$ in the interval $1 < x < \infty$ by the relation $f(x + n) = f(x)$, $0 \leqq x \leqq 1$, $(n = 1, 2, \ldots)$, and consider the expressions

$$F_p(t) = \frac{1}{10^p} \sum_{n=0}^{10^p-1} f\left(\frac{n}{10^p} + t\right), \quad (p = 1, 2, \ldots). \tag{4}$$

According to a theorem of T. J. Boks[8] to any positive number ϵ there corresponds a positive integer $p(\epsilon)$ so large that

$$\int_0^1 | F_p(t) - I | dt < \epsilon, \qquad I = \int_0^1 f(t)dt, \tag{5}$$

as soon as $p > p(\epsilon)$. We also know by (3) and (4) that $F_p(t) = f(10^p t)$. Thus, if we set

$$\int_{n\cdot 10^{-p}}^{(n+1)\cdot 10^{-p}} | F_p(t) - I | dt = \epsilon_n, \qquad (n = 0, 1, \ldots, 10^p - 1),$$

and if ϵ_k is the least of the number ϵ_n, then by (5) we obtain

$$\int_{k\cdot 10^{-p}}^{(k+1)\cdot 10^{-p}} | f(10^p t) - I | dt < \frac{\epsilon}{10^p}. \tag{6}$$

Setting $x = 10^p \cdot t$ in (6), we find

$$\int_0^1 | f(x) - I | dx < \epsilon.$$

Since this inequality holds for all positive values of ϵ, it follows that

$$f(x) = I = \int_0^1 f(t)dt$$

for all values of x save perhaps for a set of measure zero.

5. It will be shown now that $f(x)$ is equal to either 0 or 1 for all values of x save perhaps for a set of measure zero. Let us assume that $f(x) = c$ for almost all values of x, where $0 < c < 1$. Consider a net N_1, all of whose meshes are of equal length 10^{-1}, fixed onto the unit square and therefore consisting of 10^2 equal squares. Next, take a net N_2, all of whose meshes are of length 10^{-2}, dividing each square of the net N_1 in the same manner as N_1 divides the unit square. Thus, N_2 consists of 10^4 squares. The nth net N_n consists of 10^{2n} squares, each mesh being of length 10^{-n}. Continuing this indefinitely, we obtain a system of nets $\{N_n\}$. The set of all those points that lie on the side of at least one net of the system $\{N_n\}$ of nets is of superficial measure zero. The remaining points to the unit

square, and hence all points of the invariant set S, have the property that to each point P there corresponds a unique sequence of squares $S_1(P)$, $S_2(P)$, ... of the system $\{N_n\}$, each one containing P in its interior such that their areas shrink to zero. By a theorem of Lebesgue the metrical density[9] of the set S in all points of S is equal to 1 except perhaps in a set of superficial measure zero. Hence,

$$\lim_{n\to\infty} \frac{m[S\cdot S_n(P)]}{m[S_n(P)]} = 1 \tag{7}$$

for almost all points P of S.

On the other hand, it is readily seen that the transformation (T), as well as its inverse and its iterates, carry every measurable set into a measurable set of equal superficial measure. Consider again the square $S_n(P)$. By the nth iterated inverse (T^{-n}) of (T) the square $S_n(P)$ is carried into a certain rectangle $R_n(P)$ whose altitude is of length 1 and whose base is of length 10^{-2n}. Furthermore, by the remark made above the set $S\cdot S_n(P)$ goes over into a set $\Sigma_n(P)$ of equal measure. Hence,

$$\frac{m[\sum_n(P)]}{m[R_n(P)]} = \frac{m[S\cdot S_n(P)]}{m(S_n(P)]}. \tag{8}$$

Since $f(x) = c$ for almost all values of x, by Fubini's theorem mentioned in footnote,[6]

$$\frac{m[\sum_n(P)]}{m[R_n(P)]} = c. \tag{9}$$

Hence, it is evident that on comparing equations (7), (8) and (9) the assumption that $0 < c < 1$ leads to a contradiction. Consequently, $c = 0$ or $c = 1$.

Finally, since $f(x_0)$ was defined as the linear measure of that part of the invariant set S which lies on the segment $x = x_0$, $0 \leqq y \leqq 1$, by Fubini's theorem the superficial measure of S is either 0 or 1.

[1] Written while the author was a NATIONAL RESEARCH FELLOW.

[2] G. D. Birkhoff and P. A. Smith, *Jour. Mathémat.*, **7**, 9th series, 365 (1928).

[3] For an elegant treatment see E. Hopf, *Proc. Nat. Acad. Sci.*, **18**, 98 (1932).

[4] This definition differs somewhat from those given by Birkhoff, Smith and Hopf.

[5] The symbol $[\alpha]$ denotes the largest integer not greater than α.

[6] This follows from a theorem of G. Fubini, for which cf. C. Carathéodory, *Vorlesungen über reelle Funktionen*, 621–628 (1927).

[7] A different proof of this fact has also been given by Dr. S. Saks.

[8] T. J. Boks, *Rendiconti del Circolo Matematico di Palermo*, **45**, 212–213 and 226–231 (1921).

[9] For the notion and properties of metrical density cf. E. W. Hobson, *The Theory of Functions of a Real Variable*, **1**, 2nd edition, 178–182 (1921).

APPLICATION OF POINCARÉ'S SWEEPING-OUT PROCESS

BY G. C. EVANS

DEPARTMENT OF MATHEMATICS, THE RICE INSTITUTE

Communicated February 28, 1933

1. *The Conductor Potential.*—By means of Poincaré's sweeping-out process, an additive positive function of point sets $f(e)$ in a three-dimensional region T defines a similar positive function $\mu(e)$ on the boundary t of T, such that the potentials

$$U(M) = \int_T \frac{1}{MP}\, df(e_P), \quad V(M) = \int_t \frac{1}{MP}\, d\mu(e_P)$$

satisfy the relations

$$V(M) \leq U(M), \quad \nabla^2 V(M) = 0, \quad M \text{ in } T,$$
$$V(M) = U(M), \quad M \text{ not in } T + t, \quad \text{if such points exist,}$$
$$V(M) \leq U(M), \quad M \text{ on } t.$$

In particular, if T is a region containing infinity and t is a bounded (closed) set of positive capacity, the sweeping out of a mass $m = R$, initially distributed uniformly on a spherical surface of radius R, containing t in its interior, yields a positive mass μ on t, equal to the capacity of t, and the potential of this mass is the conductor potential.[1]

We restrict ourselves to such bounded sets t. It may be shown[2] that if $f(e)$ is given, and such that there is no charge in a neighborhood of t (where t is of spatial measure zero), a necessary and sufficient condition that a point Q of t be a *regular*[3] point of t is that $V(Q) = U(Q)$. In particular, $f(e)$ may designate a unit point mass at a positive distance from t.

As far as the conductor potential is concerned, the sufficiency of the condition is established immediately (without restriction on the measure of t). For $V(M)$ is lower semi-continuous. Hence if $V(Q) = 1$ for some Q in t, there is, given $\epsilon > 0$, a neighborhood of Q such that $V(M) > 1 - \epsilon$, M in this neighborhood. Consequently for M in T,

$$\lim_{M = Q} V(M) = 1,$$

and Q is a regular point.[4]

Let t be the boundary of T, t' a closed subset of t and T' the region which contains T and has t' for its boundary. It may be shown that for all points of T the potential obtained by sweeping out on t' a positive distribution $f(e)$ originally in T, at a positive distance from t, is the same as that obtained by sweeping it out on t, as a first step, and by sweeping out the resulting mass on to t', as a second step; the potential obtained

as a result of the second step is therefore dominated by that obtained by the first on all the points of t.

It is known[5] that

$$\overline{\lim_{M = Q}} \; V(M) = 1$$

for any *capable* point Q of t (an incapable point is one which is the center of some sphere such that the portion of t within that sphere is of zero capacity). Without loss of generality we may suppose that the incapable points of t are eliminated, since the remaining set is closed. We assume then that t is a *reduced* set. A reduced set has no isolated points.

2. *A Lemma on the Superior Limit of Potential at Points of the Mass.*— The following statement is to be established.

LEMMA I. *Let $\mu(e)$ be a distribution of positive mass on t of which the potential V is bounded. Let Q be a point of t, and let k be the superior limit of $V(P)$, P in t, as P tends to Q. Given $\epsilon > 0$, we can find Q_1 in t, arbitrarily close to Q, so that*

$$\overline{\lim_{\substack{M = Q_1 \\ MCT}}} \; V(M) < V(Q_1) + \epsilon < k + 2\epsilon. \tag{2.1}$$

We shall suppose the first inequality of the theorem false. Given $\epsilon > 0$, we choose a neighborhood Ω of Q small enough so that $V(P) < k + \epsilon 2^{-2}$, for P in t in this neighborhood. Let Q_1 be a point of t in Ω for which

$$k - \epsilon 2^{-3} < V(Q_1) < k + \epsilon 2^{-2}. \tag{2.2}$$

Let C_ρ be a spherical region, lying in Ω, with center Q_1 and radius ρ smal enough so that the contribution $V_{t \cdot C_\rho}(Q_1)$ to the potential $V(Q_1)$ from the mass on $t \cdot C_\rho$ satisfies the inequality (since $V(Q_1)$ is finite, lim $(\rho = 0)$ $V_{t \cdot C_\rho}(Q_1) = 0$)

$$V_{t \cdot C_\rho}(Q_1) < \epsilon 2^{-4}$$

and therefore

$$V_{t - t \cdot C_\rho}(Q_1) > k - \epsilon(2^{-3} + 2^{-4}).$$

The quantity $V_{t - t \cdot C_\rho}(P)$ is continuous at $P = Q_1$; let therefore C_δ be a second spherical neighborhood of Q_1, $\delta < \rho$, small enough so that

$$V_{t - t \cdot C_\rho}(P) > k - \epsilon(2^{-3} + 2^{-4} + 2^{-5}), \; P \text{ in } C_\delta$$

Thus, for P in $t \cdot C_\delta$,

$$V_{t \cdot C_\rho}(P) < k + \epsilon 2^{-2} - V_{t - t \cdot C_\rho}(P) < \epsilon(2^{-2} + 2^{-3} + 2^{-4} + 2^{-5}). \tag{2.3}$$

On the hypothesis that the theorem is false, we can find a sequence of points M_i in T, $i = 2,3,\ldots$, lim $(i = \infty)$ $M_i = Q_1$, such that

$$V(M_i) > V(Q_1) + \epsilon(1 - 2^{-5}). \tag{2.4}$$

Denote the distance M_iQ_1 by δ_i, taking i great enough so that $\delta_i < \delta/2$, and let Q_i be a point in t at the minimum distance, say δ_i', from M_i. There is at least one such point since t is closed. The point Q_i lies in C_δ since $Q_1Q_i \leq Q_1M_i + M_iQ_i < \delta/2 + \delta/2$; moreover lim $(i = \infty)$ $Q_i = Q_1$.

We have

$$Q_iP \leq Q_iM_i + M_iP \leq 2M_iP, \quad P \text{ in } t\cdot C_\rho; \tag{2.5}$$

in fact, $Q_iM_i = \delta_i'$, $M_iP \leq \delta_i'$. Also, for P in $t - C_\rho\cdot t$,

$$Q_iP \leq Q_iM_i + M_iP = \left(1 + \frac{Q_iM_i}{M_iP}\right) M_iP$$
$$< \left(1 + \frac{\delta_i'}{\rho - \delta/2}\right) M_iP.$$

Hence

$$Q_iP < \left(1 + \frac{\delta_i}{\rho - \delta}\right) M_iP, \quad P \text{ in } t - t\cdot C_\rho. \tag{2.6}$$

Now

$$V(M_i) = V_{t\cdot C_\rho}(M_i) + V_{t-t\cdot C_\rho}(M_i) = \int_{t\cdot C_\rho} + \int_{t-t\cdot C_\rho} \frac{1}{Q_iP}\frac{Q_iP}{M_iP}\, d\mu(e_P)$$

$$\leq 2V_{t\cdot C_\rho}(Q_i) + \left(1 + \frac{\delta_i}{\rho - \delta}\right) V_{t-t\cdot C_\rho}(Q_i)$$

$$\leq V(Q_i) + V_{t\cdot C_\rho}(Q_i) + \frac{\delta_i}{\rho - \delta} V_{t-t\cdot C_\rho}(Q_i).$$

But $V(Q_i) = V(Q_1) + V(Q_i) - V(Q_1) < V(Q_1) + k + \epsilon 2^{-2} - (k - \epsilon 2^{-3}) = V(Q_1) + \epsilon(2^{-2} + 2^{-3})$. Also $V_{t-t\cdot C_\rho}(Q_i) < k + \epsilon 2^{-2}$. And since Q_i is in C_δ we may apply (2.3). Accordingly

$$V(M_i) < V(Q_1) + \epsilon(2^{-1} + 2^{-2} + 2^{-4} + 2^{-5}) + \frac{\delta_i}{\rho - \delta}(k + \epsilon 2^{-2}).$$

Choose now i great enough so that

$$\frac{\delta_i}{\rho - \delta}(k + \epsilon 2^{-2}) < \epsilon 2^{-3}.$$

For all such values of i we have

$$V(M_i) < V(Q_1) + \epsilon(31/32).$$

This inequality, however, contradicts (2.4). The proof of the lemma is accordingly completed.

Since V is lower semi-continuous,

$$V(Q_1) \leq \lim_{P = Q_1} V(P), \quad P \text{ in } t,$$

and the latter quantity may be substituted for the former in (2.1). Incidentally, we note that the lemma applies even if t is not reduced, provided that Q is not an isolated point of t. For V is continuous in the neighborhood of Q if Q is an incapable point.

3. *Application to Conductor Potentials.*—We employ the following proposition.

LEMMA II. *Let the boundary t be a reduced bounded set of positive capacity. Let $\mu(e)$ be a distribution of positive mass on t corresponding to the conductor potential, and let*

$$V(M) = \int_t \frac{1}{MP}\, d\mu(e_P)$$

be this conductor potential. Then, for every Q in t,

$$\overline{\lim_{P = Q}}\ V(P) = 1, \quad P \text{ in } t.$$

The proof depends on Lemma I. Given $\epsilon > 0$, we can in any neighborhood of Q find Q_1 in t so that

$$\overline{\lim_{M = Q_1}}\ V(M) < V(Q_1) + \epsilon, \quad M \text{ in } T.$$

But $\overline{\lim}(M = Q_1)\ V(M) = 1$, M in T (see Art. 1). Hence $V(Q_1) > 1 - \epsilon$. Given therefore a sequence of positive values $\epsilon_1, \epsilon_2, \ldots$, tending to zero, we can find a sequence of points $P_1, P_2, \ldots$, all in t, tending to Q so that $1 \geq V(P_i) > 1 - \epsilon_i$, that is, so that

$$\lim_{i = \infty}\ V(P_i) = 1.$$

But $V(P) \leq 1$ for all P in t. Hence the conclusion of the lemma follows.

From the statement just established follows KELLOGG'S LEMMA:[6]

If t is a closed bounded set of positive capacity, it contains at least one regular point.

Without loss of generality we may assume that t consists merely of boundary points of an infinite region T, such that the complement of $T + t$ is void, since the truth of the lemma is well known for the alternative case. We may also assume that t is reduced, since if it has positive capacity it contains a reduced set of positive capacity. We assume that the theorem is not true, and therefore that $V(P) < 1$ for all P in t.

The portion t_k of t for which $V(P) \leq k < 1$, is null or closed. We may take k near enough to 1 so that t_k has positive capacity. In fact, let $k_1, k_2, \ldots$ be an increasing sequence of positive numbers tending to 1, and $e_m = t_{k_m} - t_{k_{m-1}}$. Then e_m is measurable Borel, and $t = e_1 + e_2 + \ldots$. There is at least one of the sets e_m such that $\mu(e_m) > 0$, and therefore for this m, $\mu(t_{k_m}) > 0$. Hence t_{k_m} has positive capacity,[7] and we may choose

$k = k_m$. Let t' be the reduced subset of t_k; it is not a null set and is closed.

Denote by T' the region complementary to t', and by $V(t', M)$ the conductor potential of t' at M, this being the potential of a certain distribution of positive mass on t'. Then this distribution can be chosen so that $V(t', M) \leq V(M)$ for every M in T or on t (see Art. 1). Consequently $V(t', P) \leq k < 1$ for all points of t'.

But, by Lemma II, for every point Q of t' we have

$$\overline{\lim_{P = Q}} V(t', P) = 1, \quad P \text{ in } t'.$$

Hence in any neighborhood of Q there is a point P of t' for which $V(t', P) > k$. This is a contradiction.[8]

The proof just given establishes the fact that in the neighborhood of any point Q of t (reduced, bounded), there is a point P of t for which $V(P) = 1$. For in such a neighborhood there is contained a reduced subset of t of positive capacity.

[1] C. de la Vallée Poussin, "Extension de la méthode du balayage de Poincaré, et probléme de Dirichlet," *Ann. de l'Inst. H. Poincaré*, **2**, 169–232 (1932).

[2] The necessity of this condition was stated in a theorem presented by the author to the *Amer. Math. Soc.*, Dec., 1932. Proofs of these statements will be given later in a paper of general nature.

[3] "Regular" with respect to the continuous Dirichlet problem for t. See O. D. Kellogg, *Foundations of Potential Theory*, Berlin, p. 328 (1929).

[4] Kellogg, loc. cit., p. 331.

[5] F. Vasilesco, "Sur les singularités des fonctions harmoniques," *J. de Math.*, **9**, 81–111. See p. 101 (1930).

[6] Kellogg, loc. cit., p. 337.

[7] De la Vallée Poussin, loc. cit., p. 225.

[8] A very simple substitute proof of this last theorem, independent of the citation of Art. 1, has been suggested to the author by Dr. J. J. Gergen, to whom the author is indebted for examination of the manuscirpt. Let t_k be the part of t for which

$$\frac{\lim}{M = P} V(M) \leq k < 1, M \text{ in } T,$$

Then for some k, t_k is closed, bounded and of positive capacity. Further, t' being the reduced subset of t_k,

$$V(t', M) \leq V(M), M \text{ in } t,$$

and therefore

$$\frac{\lim}{M = P} V(t', M) \leq k, P \text{ in } t', M \text{ in } T.$$

Thus the relation

$$V(t', P) \leq \frac{\lim}{M = P} V(t', M) \leq k, P \text{ in } t',$$

yields a contradiction.

GEOMETRY OF TWO-COMPONENT SPINORS

BY OSWALD VEBLEN

THE INSTITUTE FOR ADVANCED STUDY

Communicated March 14, 1933

1. The differential equations of Dirac[1] for an electron have been generalized by Weyl[2] and Fock[3] to the case of general relativity and further studied by Schouten[4] and others. The theory which results makes use of ideas which seem likely to be of considerable significance in differential geometry in general. It is the purpose of this note to set forth some of the requisite geometrical ideas in the simplest case, the case which can be handled by direct generalization of the spinors defined by van der Waerden.[5] The note is thus a sort of geometric commentary on the paper of Weyl.

At each point of an *underlying manifold* with a coördinate system $(x^1, \ldots, x^4)$ there is a *tangent space* with coördinates $(dx^1, \ldots, dx^4)$. A differential form $g_{ij}dx^i dx^j$, gives a Euclidean measure of distance, angle, etc., in each tangent space. The Riemannian geometry is the simultaneous theory of all these Euclidean spaces.

The locus

$$g_{ij}dx^i dx^j = 0 \tag{1.1}$$

is a quadratic cone through the origin. Let us confine ourselves to the relativity case, in which (1.1) is the *light-cone* and is of signature $(-, -, -, +)$. This means that there is a transformation of coördinates, $dx \longrightarrow X$, in the tangent spaces which carries the cone (1.1) into

$$-(X^1)^2 - (X^2)^2 - (X^3)^2 + (X^4)^2 = 0. \tag{1.2}$$

Let us write the equations of this transformation in the form,

$$X^a = g^a_i dx^i. \tag{1.3}$$

We adopt the convention that the letters $a, \ldots, h$ used as indices refer to the coördinate system X in the tangent space, and $i, \ldots, z$ used as indices refer to the coördinates of the underlying space or of the differentials $dx^1, \ldots, dx^4$. We shall also use as far as conveniently possible, the convention of Schouten that a geometrical object shall be indicated in all coördinate systems by the same carrier letter, different sorts of coördinate systems being indicated by different alphabets or portions of alphabets used as indices. Thus the quadratic form in (1.2) is

$$g_{ab}X^a X^b.$$

Also we have

$$g_{ab}g^a_i g^b_j = g_{ij},$$

where

$$(g_{ab}) = \begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1. \end{matrix} \tag{1.4}$$

We make the definition

$$g^{ia} = g^{ai} = g^{ij}g^a_j \tag{1.5}$$

and

$$g^i_a = g_{ab}g^{ib}, \tag{1.6}$$

where, as usual

$$g^{ik}g_{jk} = \delta^i_j.$$

It then follows that

$$g^i_j = g^i_a g^a_j = \delta^i_j \text{ and } g^i_a g^b_i = g^b_a = \delta^a_b. \tag{1.7}$$

Having the existence of the quantities g^a_j in one coördinate system x we are free to assume a transformation law for them. We assume that g^1_i, g^2_i, g^3_i, g^4_i are four covariant vectors. It then follows that the components of these vectors in any other coördinate system will transform the equation (1.1) in the new coördinate system into (1.2). If we assume also that g^i_1, g^i_2, g^i_3, g^i_4 are four contravariant vectors, then (1.5), (1.6) and (1.7) hold in all coördinate systems.

From the fact that g^1_i, g^2_i, g^3_i, g^4_i are covariant vectors, it follows that X^1, X^2, X^3, X^4 are unaltered by any transformation of coördinates x in the underlying space. We therefore call them *scalar coördinates* for the tangent Euclidean space. They can be subjected to an arbitrary Lorentz transformation

$$\overline{X}^a = L^a_b X^b \tag{1.8}$$

without altering anything that has been said, provided that we replace (1.3) by

$$\overline{X}^a = \bar{g}^a_j dx^j$$

where

$$\bar{g}^a_j = L^a_b g^b_j.$$

With respect to transformations of coördinates in the underlying space the quantities L^a_b are 16 scalars.

The quantities g_{ai}, g^a_i, g^i_a, etc., are components of the metric tensor with respect to a in the scalar coördinate system X, and with respect to i in the coördinate system dx.

2. The cone (1.1) is a system of straight lines through its vertex. The structure of this system of lines may be made evident as follows. Consider any 3-space S_3 in the tangent 4-space but not passing through the origin (point of contact). Each line of the cone cuts S_3 in one and but

one point and the set of all these points constitutes a quadric surface in S_3. Hence the relationships among themselves of the lines of the cone are the same as those among the points of an ordinary quadric in a 3-space.

The property of a quadric which we wish to use is that it contains two one-parameter systems of straight lines, the generators of the quadric. Any line of one system cuts all the lines of the other system, and no two lines of the same system intersect. If λ is a parameter which designates a line of the first system and μ a parameter which designates a line of the second system then the pair of numbers (λ, μ) designates a unique point on the quadric. Inversely any point of the quadric determines a pair of parameter values (λ, μ) since there is one line of each system through such a point.

Going back to the cone (1.1), each point of the quadric is the trace in S_3 of a line of the cone and each line on the quadric is the trace of a plane on the cone. Hence there are two systems of planes on the cone such that two planes of the same system have no point in common except the vertex and such that each plane of one system intersects each plane of the other system in a line. The planes of one system can be designated by the values λ and those of the other system by μ and the pair of values (λ, μ) will determine a line of the cone.

In the case which we are studying, when the equation of the cone is reducible to (1.2), the quadric in S_3 is real but the straight lines on it are imaginary. The quadric is projectively equivalent to an ordinary sphere. Thus the structure of the system of light lines through the vertex of (1.1) is like that of the points on a sphere.

The properties which we have been inferring from theorems of geometry can also be obtained by an elementary algebraic argument. Suppose we set

$$\begin{aligned} \frac{X^1 + iX^2}{\sqrt{2}} &= X^{14} & \frac{X^3 + X^4}{\sqrt{2}} &= X^{13} \\ \frac{X^1 - iX^2}{\sqrt{2}} &= X^{23} & \frac{X^3 - X^4}{\sqrt{2}} &= -X^{24}. \end{aligned} \tag{2.1}$$

Then

$$2\begin{vmatrix} X^{13} & X^{14} \\ X^{23} & X^{24} \end{vmatrix} = -(X^1)^2 - (X^2)^2 - (X^3)^2 + (X^4)^2. \tag{2.2}$$

If this determinant is to vanish there must exist variables $\psi^1, \psi^2, \psi^3, \psi^4$, such that

$$\begin{aligned} X^{13} &= \psi^1\psi^3 & X^{14} &= \psi^1\psi^4 \\ X^{23} &= \psi^2\psi^3 & X^{24} &= \psi^2\psi^4. \end{aligned} \tag{2.3}$$

Solving the equations (2.1) we obtain

$$\begin{aligned} X^1 &= \frac{X^{14} + X^{23}}{\sqrt{2}} = \frac{\psi^1\psi^4 + \psi^2\psi^3}{\sqrt{2}} \\ X^2 &= \frac{X^{14} - X^{23}}{\sqrt{-2}} = \frac{\psi^1\psi^4 - \psi^2\psi^3}{\sqrt{-2}} \\ X^3 &= \frac{X^{13} - X^{24}}{\sqrt{2}} = \frac{\psi^1\psi^3 - \psi^2\psi^4}{\sqrt{2}} \\ X^4 &= \frac{X^{23} + X^{24}}{\sqrt{2}} = \frac{\psi^1\psi^3 + \psi^2\psi^4}{\sqrt{2}} \end{aligned} \tag{2.4}$$

as a parameter representation of the points on the cone (1.1).

If ψ^3 and ψ^4 are held fixed and ψ^1 and ψ^2 regarded as variable parameters these equations say that the point (X^1, X^2, X^3, X^4) describes the plane through the vertex and the points $(\psi^4, -i\psi^4, \psi^3, \psi^3)$ and $(\psi^3, i\psi^3, -\psi^4, \psi^4)$. If ψ^3 and ψ^4 are multiplied through by the same factor, this plane is unaltered. If, however, the ratio ψ^3/ψ^4 is changed the plane is changed and there is one such plane for each value of this ratio. Moreover it is easy to infer from the equations (2.4) that no two of the planes so obtained have any part in common except (0, 0, 0, 0).

In like manner we see that there is another system of planes on the cone, one for each value of ψ^1/ψ^2. Two planes, one from the first system determined by (ψ^3, ψ^4), and one from the second determined by (ψ^1, ψ^2) have a line in common. This line is given by (2.4) with the understanding that the coördinates ψ^1, ψ^2 and ψ^3, ψ^4 may be multiplied through by an arbitrary factor. Thus (2.4) is a parameter representation of the points on the light-cone.

From the geometry of the quadric we know that any linear transformation of the quadric into itself brings about a projective transformation of the generating lines. With this suggestion we subject ψ^1 and ψ^2 to a linear transformation,

$$\bar{\psi}^A = T^A_B\psi^B. \qquad (A, B = 1, 2). \tag{2.5}$$

Substituting this in (2.4) we see that the X's undergo a linear transformation whose coefficients can easily be calculated. The determinant in (2.2) is evidently multiplied by the square of the determinant

$$T = |T^A_B| \tag{2.6}$$

when the transformation (2.5) is carried out. Hence the transformations (2.5) induce transformations of the light-cone into itself. Similarly for linear transformations of ψ^3 and ψ^4.

Instead of regarding (2.5) as a geometric transformation which permutes one system of generating planes of the light cone among themselves, we may regard it as a transformation of the parameters ψ^1, ψ^2 into new parameters, i.e., as a change of the reference system for the light-planes. Because of the physical applications we will refer to ψ^1, ψ^2 (and also ψ^3, ψ^4) as *spin coördinates*. Thus (2.5) can be regarded as a transformation of spin coördinates which brings about a change of the spin-frame of reference.

We shall use the conventions of tensor algebra to indicate the well-known transformation laws that depend on (2.5), for example, the covariant transformation

$$\bar{\psi}_A = t_A^B \psi_B \qquad (A, B = 1, 2) \tag{2.7}$$

where

$$t_B^A T_C^B = \delta_C^A. \tag{2.8}$$

Let us define ϵ_{AB} as having the components

$$\epsilon_{11} = \epsilon_{22} = 0, \quad \epsilon_{12} = -\epsilon_{21} = 1.$$

Then, evidently, the components ϵ_{AB} are the same in all spin-frames if we assume the law of transformation of a relative covariant tensor of weight -1

$$\epsilon_{CD} = \frac{1}{t} \epsilon_{AB} t_C^A t_D^B.$$

The weight is the exponent of the determinant,

$$t = |t_B^A|. \tag{2.9}$$

In like manner we may define

$$\epsilon^{11} = \epsilon^{22} = 0, \qquad \epsilon^{12} = -\epsilon^{21} = 1$$

and find

$$\epsilon^{CD} = t\epsilon^{AB} T_A^C T_B^D$$

the transformation law of a relative contravariant tensor of weight $+1$.

We shall use the two ϵ's to raise and lower indices as follows

$$\epsilon_{BA}\psi^B = \psi_A \tag{2.10}$$

or

$$\psi_1 = -\psi^2, \quad \psi_2 = \psi^1$$

$$\epsilon^{AB}\psi_B = \psi^A. \tag{2.11}$$

Notice that in lowering indices we sum with respect to the first index of

ϵ_{AB} while in raising indices we sum with respect to the second index of ϵ^{AB}. This corresponds to the identities

$$-\epsilon_{AB}\epsilon^{CA} = \epsilon_{AB}\epsilon^{AC} = \delta^C_B. \tag{2.12}$$

It must be remembered that raising indices increases the weight by $+1$ and lowering them decreases it by -1. Thus if we assume ψ^A to be of weight $+1/2$ then ψ_A must be of weight $-1/2$.

We actually do assume that the ψ^A which enter the formulas (2.4) for the parameterization of the light-cone are of weight $+1/2$ because this means that a change of spin-frame causes the transformation

$$\bar{\psi}^A = t^{1/2}T^A_B\psi^B \tag{2.13}$$

which is of determinant 1 and therefore does not change the value of the left-hand member of (2.2).

3. The equations (2.1) can be written in abbreviated form as

$$X^{AP} = g^{AP}_a X^a \tag{3.1}$$

in which A takes on the values 1 and 2, P the values 3 and 4 and a the values 1, 2, 3, 4. Throughout this paper we shall use the early capital letters, previous to P, as indices to denote the values 1 and 2, and the later capital letters, from P onward, as indices to denote 3 and 4.

The coefficients of (3.1) for the four values of a constitute four 2-row Hermitian matrices

$$g^{\cdot\cdot}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, g^{\cdot\cdot}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}0 & i\\ -i & 0\end{pmatrix}, g^{\cdot\cdot}_3 = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, g^{\cdot\cdot}_4 = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}. \tag{3.2}$$

The equations (2.4) which are inverse to (3.1) may be written,

$$X^a = g^a_{AP}X^{AP} \tag{3.3}$$

where

$$g^1_{\cdot\cdot} = \frac{1}{\sqrt{2}}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, g^2_{\cdot\cdot} = \frac{1}{\sqrt{2}}\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}, g^3_{\cdot\cdot} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, g^4_{\cdot\cdot} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}.$$

Since (3.1) and (3.3) are inverse, we have

$$g^{AP}_a g^b_{AP} = \delta^b_a \tag{3.4}$$

and

$$g^{AP}_a g^a_{BQ} = \delta^A_B\delta^P_Q. \tag{3.5}$$

Let us now apply the conventions agreed upon in the last section to raising and lowering the indices. Thus

$$g^P_{Aa} = \epsilon_{BA}g^{BP}_a.$$

This gives four matrices

$$g_{:1} = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad g_{:2} = \frac{1}{\sqrt{2}}\begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix}, \quad g_{:3} = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad g_{:4} = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{3.6}$$

If we calculate the other possible arrangements of indices we find that the operation of raising and lowering indices as defined in the sections above is consistent with the relations (3.4) and (3.5).

From the equation,

$$g^b_a g^a_{AP} = g^b_{AP}$$

it is evident that the g's which appear in this section are all components of the metric tensor, some in the scalar and some in the spin-frames of reference. For example, we have

$$g_{APBQ} = \epsilon_{AB}\epsilon_{PQ}. \tag{3.7}$$

Also we have

$$\delta^A_B g^{ab} = g^{Aa}_P g^{Pb}_B + g^{Pa}_B g^{Ab}_P, \tag{3.8}$$

as may readily be verified.

The components of the metric tensor in the scalar and the spin coördinates are all constants, but this is not true of components, partly in the dx coördinates and partly in the spin coördinates. For example the components in the left-hand member of the equation

$$g^{Ai}_P = g^{Aa}_P g^i_a$$

are not, in general, constants. Nevertheless we have

$$\delta^A_B g^{ij} = g^{Ai}_P g^{Pj}_B + g^{Aj}_P g^{Pi}_B$$

or, in a matrix notation,

$$1.g^{ij} = g^i g^j + g^j g^i.$$

4. The condition that the coördinates X^1 shall be real gives by (2.1) that

$$X^{13} \text{ and } X^{24} \text{ are real}$$

and

$$X^{14} = X^{23*}$$

where we are using the * to indicate conjugate imaginaries.

From (2.3) it follows that

$$\begin{aligned} \psi^{3*} &= k\psi^1 \\ \psi^{4*} &= k\psi^2 \end{aligned} \tag{4.1}$$

where k is real. The effect of changing the value of k is merely to shift the point X along the same line of the light-cone. Hence no generality is lost by assuming $k = 1$ as we now shall do.

Hence the points X on the light-cone are given by

$$\begin{aligned} X^1 &= \frac{\psi^1\psi^{2*} + \psi^2\psi^{1*}}{\sqrt{2}} \\ X^2 &= \frac{\psi^1\psi^{2*} - \psi^2\psi^{1*}}{\sqrt{-2}} \\ X^3 &= \frac{\psi^1\psi^{1*} - \psi^2\psi^{2*}}{\sqrt{2}} \\ X^4 &= \frac{\psi^1\psi^{1*} + \psi^2\psi^{2*}}{\sqrt{2}}. \end{aligned} \tag{4.2}$$

From the form of (2.4) it is clear that if ψ^1 and ψ^2 are multiplied by a number c and ψ^3 and ψ^4 by $1/c$ the point X represented by (2.4) is unchanged. But if the relation (4.1) is to be undisturbed, we must have

$$cc^* = 1.$$

Hence the point X which is represented by (4.2) is unchanged if we multiply ψ^1 and ψ^2 through by $\exp(i\theta)$ where θ is real. On the other hand a different point on the same generating line of the cone is represented by these equations if we multiply ψ^1 and ψ^2 by a constant whose absolute value is not 1. Each value of the ratio ψ^1/ψ^2 determines a line of the cone.

The parameters ψ like the coördinates X, are unaltered by changes of coördinates in the underlying space-time. But in a given coördinate system x they have a degree of indetermination which is expressed by writing

$$\begin{aligned} \psi^1 &= e^{ix^0}f^1 \\ \psi^2 &= e^{ix^0}f^2 \end{aligned} \tag{4.3}$$

where x^0 is an arbitrary real parameter.

If f^1 and f^2 are taken to be complex functions of the real variables x^1, ..., x^4 these equations determine a unique real point on the light-cone at each point x. This point is the same for all values of x^0. If we make a transformation

$$\bar{x}^0 = x^0 + \log \rho(x^1, x^2, x^3, x^4) \tag{4.4}$$

where $\log \rho$ is real then f^1 and f^2 are multiplied by $1/\rho$ but the same point on the same light-cone is represented by (4.3). Hence a pair of projective scalars[6] ψ^1 and ψ^2 determine a unique point of the light-cone at each point of space-time.

We may, indeed, make the more general transformation

$$\bar{x}^0 = x^0 + \int \lambda_j dx^j \tag{4.5}$$

instead of (4.4) which changes (4.3) into

$$\psi^A = f^A \exp.\ (i\bar{x}^0 - i\int \lambda_j dx^j) \tag{4.6}$$

and introduces a parametrization of the light-cone which depends on the path of integration, without changing the point on the light-cone represented by (4.2).

5. Let us now require of our transformations of spin-frame that the reality condition of §4 shall be undisturbed. This means that T^P_Q must be related to T^A_B by the equation

$$T^{A+2}_{B+2} = T^{A*}_{B}. \tag{5.1}$$

Hence the determinant of the transformation of ψ^3 and ψ^4 is T^*.

As indicated at the end of §2 we subject the spin parameters to the laws of transformation

$$\bar{\psi}^A = t^{1/2} T^A_B \psi^B \tag{5.2}$$

$$\bar{\psi}^P = t^{*1/2} T^P_Q \psi^Q \tag{5.3}$$

and, consistently with this,

$$\bar{X}^{AP} = (t\,t^*)^{1/2} X^{BQ} T^A_B T^P_Q, \tag{5.4}$$

which leaves the quadratic form

$$\begin{vmatrix} X^{13} & X^{14} \\ X^{23} & X^{24} \end{vmatrix}$$

unaltered. Hence

$$L^a_b = (t\,t^*)^{1/2} g^a_{AP} g^{BQ}_b T^A_B T^P_Q \tag{5.5}$$

is a Lorentz transformation.

6. We now have before us a new class of geometrical objects with the following characteristics. Consider two functions,

$$\begin{aligned} \psi^1 &= e^{ix^0} f^1(x^1, x^2, x^3, x^4) \\ \psi^2 &= e^{ix^0} f^2(x^1, x^2, x^3, x^4). \end{aligned} \tag{6.1}$$

If there is fixed a coördinate system, a gauge and a spin-frame, then the equations (6.1) together with (4.2) determine a definite point on each light-cone. Under a transformation of coördinates or of gauge ψ^1 and ψ^2 behave like projective scalars of index i, that is, the transformation is effected by substituting

$$\begin{aligned} x^0 &= \bar{x}^0 - \log \rho(\bar{x}) \\ x^i &= x^i(\bar{x}) \end{aligned} \tag{6.2}$$

in (6.1). We may also allow the non-holonomic gauge transformation (4.5). A change of the spin-frame requires that ψ^1 and ψ^2 be replaced by $\bar{\psi}^1$ and $\bar{\psi}^2$ according to (5.2). The geometrical object having components ψ^1 and ψ^2 subject to all these conditions, we will call a contravariant *spinor* of the first order of weight $1/2$. Spinors of other orders and weights are defined by replacing (5.2) by the suitable linear relations after the fashion of tensor algebra.

7. In order to have a theory of covariant differentiation of spinors we have only to introduce the analogue of an affine connection. This is a geometric object whose 20 components

$$\Lambda^A_{B\alpha} \quad (A, B = 1, 2;\ \alpha = 0, 1, 2, 3, 4)$$

are functions of x^1, x^2, x^3, x^4, alone and under changes of spin-frame obey the transformation law

$$\bar{\Lambda}^C_{D\alpha} = \left(\Lambda^A_{B\alpha} t^B_D + \frac{\partial t^A_D}{\partial x^\alpha}\right) T^C_A \tag{7.1}$$

whereas under transformation of coördinates and gauge they behave like components of a projective covariant vector. We shall call this a *spinor connection of the first kind.*

It then follows by the same calculation as in tensor analysis[7] that if ψ^A is a spinor of weight N

$$\frac{\partial \psi^A}{\partial x^\alpha} + \Lambda^A_{B\alpha}\psi^B - N\,\Lambda^B_{B\alpha}\psi^A = \psi^A_{,\alpha} \tag{7.2}$$

are the components of a geometric object which transform like those of a spinor of weight N with respect to the index A and like those of a projective tensor with respect to α. That is to say if we change coördinates, gauge and spin-frame all at once we have

$$\bar{\psi}^A_{,\alpha} = t^{1/2}\psi^B_{,\beta} T^A_B \frac{\partial x^\beta}{\partial \bar{x}^\alpha}. \tag{7.3}$$

We call $\psi^A_{,\alpha}$ the covariant derivative of ψ^A.

Covariant differentiation of spinors of arbitrary order and weight is defined analogously by the same formulas as in tensor analysis.[7] For example, if x^{AP} is a spinor with the law of transformation (5.4)

$$X^{AP}_{,\alpha} = \frac{\partial X^{AP}}{\partial x^\alpha} + \Lambda^A_{B\alpha}X^{BP} + \Lambda^P_{Q\alpha}X^{AQ} - \frac{1}{2}\Lambda^B_{B\alpha}X^{AP} - \frac{1}{2}\Lambda^Q_{Q\alpha}X^{AP}. \tag{7.4}$$

Our reality conditions make it necessary, of course, that

$$\Lambda^{A+2}_{B+2\alpha} = \Lambda^{A*}_{B\alpha}. \tag{7.5}$$

Hence

$$\Lambda^{P}_{P\alpha} = \Lambda^{A*}_{A\alpha}. \tag{7.6}$$

If $\Lambda^{A}_{B\alpha}$ are components of a spinor connection of the first kind the functions

$$\Gamma^{A}_{B\alpha} = \Lambda^{A}_{B\alpha} - \frac{1}{2}\Lambda^{C}_{C\alpha}\delta^{A}_{B} \tag{7.7}$$

are the components of a geometric object which we shall call a *spinor connection of the second kind*. Its law of transformation under changes of spin-frame is

$$\bar{\Gamma}^{A}_{B\alpha} = \left(\Gamma^{C}_{D\alpha}t^{D}_{B} + \frac{\partial t^{C}_{B}}{\partial x^{\alpha}}\right) T^{A}_{C} - \frac{1}{2}\frac{\partial \log t}{\partial x^{\alpha}}\delta^{A}_{B} \tag{7.8}$$

and it satisfies the invariant condition

$$\Gamma^{A}_{A\alpha} = 0. \tag{7.9}$$

In case $N = {}^1/_2$, that is in case ψ^A is a spinor of weight ${}^1/_2$ and index i, (7.2) may be written

$$\frac{\partial \psi^{A}}{\partial x^{\alpha}} + \Gamma^{A}_{B\alpha}\psi^{B} = \psi^{A}_{,\alpha} \tag{7.10}$$

8. The formula (7.10) for covariant differentiation of a spinor can be used to obtain a displacement formula as follows The equations

$$\psi^{A}_{,\alpha} - \psi^{A}_{,0}\varphi_{\alpha} = 0 \tag{8.1}$$

where φ_α is an arbitrary projective vector such that

$$\varphi_0 = 1, \tag{8.2}$$

are invariant under transformations of coördinates, gauge and spin-frame. Also they are satisfied identically for $\alpha = 0$. Hence the equations

$$(\psi^{A}, i - \psi^{A}_{,0}\varphi_{i})\frac{dx^{i}}{dt} = 0 \tag{8.3}$$

are invariant under all three types of transformation.

It is sufficient for our present purposes if we limit attention to the case in which[8]

$$\Gamma^{A}_{B0} = 0. \tag{8.4}$$

Under these conditions the components Γ^{A}_{Bi} are unaltered by gauge transformations and respond to transformations of coördinates like the com-

ponents of a covariant affine vector. Substituting (7.10) in (8.3) we now obtain

$$\frac{d\psi^A}{dt} + \Gamma^A_{Bj}\psi^B \frac{dx^j}{dt} = i\varphi_j \frac{dx^j}{dt}\psi^A. \tag{8.5}$$

This is the formula for the displacement of the ψ's which enter in the parameterization (4.3) of the light-cone along a curve

$$x^i = x^i(t).$$

Applying this and the corresponding formula for the conjugate imaginary ψ's to

$$X^{AP} = \psi^A\psi^P$$

we find the following displacement formula,

$$\frac{dX^{AP}}{dt} + \Gamma^A_{Bj}X^{BP}\frac{dx^j}{dt} + \Gamma^P_{Qj}X^{AQ}\frac{dx^j}{dt} = 0, \tag{8.6}$$

which in the scalar coördinates of §1 becomes

$$\frac{dX^a}{dt} + \Gamma^a_{bj}X^b\frac{dx^j}{dt} = 0, \tag{8.7}$$

where

$$\Gamma^a_{bj} = g^a_{AP}\Gamma^A_{Bj}g^{BP}_b + g^a_{AP}\Gamma^P_{Qj}g^{AQ}_b. \tag{8.8}$$

These functions satisfy the condition,

$$\Gamma^a_{aj} = 0 \tag{8.9}$$

in consequence of (7.9).

9. The equation (8.7) is the formula for an affine displacement and we can in fact identify this displacement with the displacement by infinitesimal parallelism determined by the gravitational tensor g_{ij}. The latter is given by

$$\frac{dX^i}{dt} + \Gamma^i_{jk}X^j\frac{dx^k}{dt} = 0 \tag{9.1}$$

in the coördinates x, where $X^i = g^i_a X^a$ and Γ^i_{jk} are the Christoffel symbols of the second kind formed from g_{ij}. Under the transformation (1.2) we find

$$\Gamma^a_{bj} = \left(\Gamma^i_{kj}g^a_i - \frac{\partial g^a_k}{\partial x^j}\right)g^k_b \tag{9.2}$$

and also, as is well known, that these functions satisfy the condition (8.9).

The equation (8.8) now yields

$$g^{AP}_a\Gamma^a_{bj}g^b_{BQ} = \Gamma^A_{Bj}\delta^P_Q + \Gamma^P_{Qj}\delta^A_B. \tag{9.3}$$

Setting $A = B$ and $P = Q$ and summing we obtain

$$\Gamma^A_{Aj} = -\Gamma^P_{Pj} = \lambda_j. \tag{9.4}$$

We obtain

$$\Gamma^A_{Bj} - \delta^A_B\lambda_j = {}^1/_2\Gamma^a_{bj}g_a^{AP}g^b_{BP} \tag{9.5}$$

by setting $P = Q$ in (9.3) and summing. This equation determines Γ^A_{Bj} uniquely if in accordance with (7.9) we assume $\lambda_j = 0$. Similarly we obtain

$$\Gamma^P_{Qj} = {}^1/_2\Gamma^a_{bj}g_a^{AP}g^b_{AQ}. \tag{9.6}$$

Thus a displacement (9.1) determines a pair of conjugate spin displacements

$$\nabla^{PB}_{AC}\psi^C = 0, \quad \nabla^{AQ}_{PR}\psi^R = 0,$$

where

$$\nabla^{PB}_{AC} = g_A^{Pj}\left(\delta^B_C\frac{\partial}{\partial x^j} + \Gamma^B_{Cj} - i\varphi_j\delta^B_C\right)$$

and

$$\nabla^{AQ}_{PR} = g_P^{Aj}\left(\delta^Q_R\frac{\partial}{\partial x^j} + \Gamma^Q_{Rj} + i\varphi_j\delta^Q_R\right).$$

The functions φ_j are completely arbitrary.

[1] P. A. M. Dirac, "The Quantum Theory of the Electron," *Proc. Roy. Soc.*, **A 117**, 610 (1928).

[2] H. Weyl, "Elektron und Gravitation," *Zeitsch. Phys.*, **56**, 330 (1929).

[3] V. Fock, "Geometrisierung der Diracschen Theorie des Elektrons," *Ibid.*, **57**, 261 (1929).

[4] J. A. Schouten, "Dirac Equations in General Relativity," *Jour. Math. and Phys.*, **10**, 239 (1931).

[5] B. L. van der Waerden, "Spinoranalyse," *Göttingen Nachrichten*, 100 (1929).

[6] For the definition of projective scalars and of projective tensors in general, see O. Veblen, "Projektive Relativitätstheorie," *Ergebnisse der Math.*, **2**, Berlin, 1933.

[7] See for example O. Veblen, *Invariants of Quadratic Differential Forms*, Cambridge, pp. 37, 38 (1927).

[8] The introduction of this simplifying assumption at this point was suggested by Mr. J. L. Vanderslice.

A CORRECTION

BY GEORGE D. BIRKHOFF

DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY

Communicated April 4, 1933

In my paper "Some Remarks Concerning Schrödinger's Wave Equation" (these PROCEEDINGS, March, 1933), the equation (6) requires a term $2Q$ within the parentheses where Q is related to L_1 just as P is to L_0. The invariant integral (9) should be written

$$I = \int_{V(\tau)} Gv_0^2 dx_1 \ldots dx_n$$

where G is any solution of the linear partial differential equation in G *for the given phase function* S:

$$\sum_{i=1}^{n} \left(\frac{\partial P}{\partial y_i} \frac{\partial G}{\partial x_i} + \frac{\partial^2 P}{\partial x_i \partial y_i} G \right) - QG = 0.$$

The proof is only modified in detail. In the special case of the Schrödinger equation with rectangular coördinates, $G = 1$ furnishes a solution while $Q \equiv 0$.

TWO OMISSIONS

In my article with T. W. Vaughan in the October, 1932, issue of the PROCEEDINGS there should have been inserted on page 611 between the next to the last and the last paragraphs a caption, namely:

Asterorbis rooki Vaughan and Cole, n. sp.
Plate 1, figs. 1–6

And on page 614 there should have appeared:

Pseudorbitoides israelskii Vaughan and Cole, n. sp.
Plate 2, figs. 1–7

placed between the three captions which head that page and the first main paragraph of the text.

W. STORRS COLE